2016

Representation and Analysis of Multi-Modal, Nonuniform Time Series Data: An Application to Survival Prognosis of Oncology Patients in an Outpatient Setting

Jennifer Winikus
Michigan Technological University, jawiniku@mtu.edu

REPRESENTATION AND ANALYSIS OF MULTI-MODAL, NONUNIFORM TIME SERIES DATA: AN APPLICATION TO SURVIVAL PROGNOSIS OF ONCOLOGY PATIENTS IN AN OUTPATIENT SETTING

By

Jennifer Winikus

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In Computer Engineering

MICHIGAN TECHNOLOGICAL UNIVERSITY

2016

This dissertation has been approved in partial fulfillment of the requirements for the Degree of DOCTOR OF PHILOSOPHY in Computer Engineering.

Department of Electrical and Computer Engineering

Dissertation Co-advisor:	*Dr. Laura E. Brown*
Dissertation Co-advisor:	*Dr. Timothy C. Havens*
Committee Member:	*Dr. Nilufer Onder*
Committee Member:	*Dr. Jinshan Tang*
Department Chair:	*Dr. Daniel Fuhrmann*

Dedication

To those who were lost along the way. A special thanks to Dr. Jalal Baghdadchi who inspired my interest to pursue graduate school beyond Alfred University along with the interest in genetic algorithms, and to Dr. Glen Archer who while at Michigan Technological University helped shape the education into a career that I love and makes a difference.

Contents

List of Figures

List of Tables

Acknowledgments

Thank you to EMOL Health of Clawson, MI for providing the data used.

Portage, a high performance computing cluster at Michigan Technological University, was used in obtaining selected results presented in this dissertation.

List of Abbreviations

ALB	Albumin
ANN	Artificial Neural Network
APCA	Adaptive Piecewise Constant Approximation
BN	Bayesian Network
CPT	Conditional Probability Table
DAG	Directed Acyclic Graph
DFT	Discrete Fourier Transform
DS	Data Set
DT	Decision Tree
DTFT	Discrete Time Fourier Transform
DTW	Dynamic Time Warping
ECG	Electrocardiogram
EM	Expectation Maximization
FT	Fourier Transform
FFT	Fast Fourier Transform
GA	Genetic Algorithm
GP	Gaussian Process
HGB	Hemoglobin

IEEE	Institute of Electrical and Electronics Engineers
IFT	Inverse Fourier Transform
IDFT	Inverse Discrete Fourier Transform
LOS	Length of Survival
NB	Naïve Bayes
NPI	Nottingham Prognostic Index
NUFFT	Nonuniform Fourier Transform
MAP	Maximum a Posteroiori
PAA	Piecewise Aggregate Approximation
PLR	Piecewise Linear Representation
PSD	Positive Semi-Definite
RF	Random Forest
RV	Random Variable
SAX	Symbolic Aggregate Approximation
SSL	Semi-supervised Learning
SVM	Support Vector Machines
SVR	Support Vector Regression
UCI	University of California Irvine
UCR	University of California Riverside
WSK	Whittaker-Kotel'nikov-Shannon
WT	Weight

Abstract

The representation of nonuniform, multi-modal, time-limited time series data is complex and explored through the use of discrete representation, dimensionality reduction with segmentation based techniques, and with behavioral representation approaches. These explorations are done with a focus on an outpatient oncology setting with the classification and regression analysis being used for length of survival prognosis. Each decision of representation and analysis is not independent, with implications of each decision in method for how the data is represented and then which analysis technique is used. One unique aspect of the work is the use of outpatient clinical data for patients, which was explored initially through discrete sampling and behavioral representation. The length of survival was evaluated with both classification and regression methods initially. The first conclusion determined that including more discrete samples in the model showed no statistical benefit and the addition of behavioral approaches did improve the prognostic accuracy.

From this result, the adaption of Piecewise Aggregate Approximation was made to accommodate the multi-modal time series data of the outpatient clinical data, and evaluated with the regression methodologies. This representation approach demonstrated promise due to the simplicity but had decreased performance in the length of survival prognosis compared with behavioral representation and discrete samples

approach. A solution was a new representation approach made which incorporates a genetic algorithm to select the window boundaries of the Piecewise Aggregate Approximation method. This selection is based on the fraction of the Piecewise Aggregate Approximation windows that contain values other than zero. The new representation improved the performance in some cases by a 20% reduction in median relative error.

Chapter 1

Introduction

Data is gathered, represented, and used in many different ways everyday. The approaches to representation and how they are used is not independent, and can be complicated by features of the process in which the data is gathered. The work looks at the representation and analysis in the application of outpatient oncology prognosis.

With every passing second, measurements and observations are being collected in many different applications around the world. As the observations are recorded, not all of them are being made with uniform frequency. One area where nonuniform observation is prevalent is outpatient medicine. In outpatient medicine, there is nonuniformity in the observations across time, type (e.g., patient's weight, height, blood tests, etc.), number (how many observations are made), and patients in a population.

Over all the observations, while a patient is alive, the values are realistically continuous but are taken as discrete observations. The observations are limited to a starting observation time and then either a time corresponding to the ending of care being received or death, making the data time-limited. Based on these properties the data set which exists in the domain of outpatient medical care is multi-modal, nonuniform, time-limited time series.

Medical data is often used to makes decisions and predictions, such as oncology survival prognosis or prescribe medications to treat simple ailments of low iron. The individual challenges of data representation and prognosis problem class (e.g, regression, binary classification, multi-class classification) are not independent of each other, with influences to prognostic accuracy resulting from modification to either aspect. The nature of oncology prognosis with outpatient medical data provides the foundation scenario for the research to be conducted. With this scenario it can be decomposed into two components: the representation of the data, and the methods and type of prognosis. In this work, prognosis will be considered as a classification problem and then regression problem.

In general, the data set of focus is a set of independent samples, with each sample consisting of observations for multiple types of measurement made over time. Each observation is representative of an attribute (also known as feature or of a sensor entity); with the observation properties (e.g., mode, measurement approaches, etc.)

independent from all other attributes. Observations are presented initially based on the attribute it belongs to, with one of three mode types considered: categorical, numerical valued, or time series. Categorical observations are when a qualitative observation is needed as a measurement (e.g., the gender association or if a patient is still alive). Numerical valued data can be viewed as a measurement over a specified duration, e.g., number of treatments in a given period or the amount of radiation exposure. When the mode considered is time series, the observations are made with respect to time (e.g., the collection of vitals or results of blood tests). For each of the attributes, the mode is held consistent in type but independent with respect to the number of observations for all other attributes and samples. This independence is one aspect of nonuniformity considered.

For any observation in the medical domain there are two additional time considerations. The first consideration is that the maximum time duration is limited. For many areas, there is a limit to how much history is known and then the observation time will end at some point (e.g., from leaving the care of the physician or dying). Within the entire collection of observations, since there are multiple attributes, the duration for each also can display nonuniform behavior which must be addressed in the representation of the data set. While the observations are time-limited, there is the second consideration in the determination of how much of the observed time should be incorporated when determining a prognosis, more time does not necessarily produce better results [6].

The distribution of observations over time within the attribute may vary substantially, there could be periods with more or less observations relative to the rest of the time series. Clusters of observations with scattered observations elsewhere that does not have any similarity to a traditional sampling distribution can be observed. Potentially the absence of observations may extend to an entire attribute. The cases on zero observation presents one of the challenges in the representation approach to be utilized.

To consider representing similar data, especially with time series data, dimensionality reduction has been used [7]. Some methods to represent the data with a reduced dimensionality include rule based [8, 9, 10], interpolation sampling [11], and artificial neural networks [12]. Other approaches to achieve dimensionality reduction to represent the data take a piecewise approach, splitting the duration of the observation time which is to be considered into a piecewise set. In the piecewise methods, the segments (or windows, or frames) are mostly done with uniform widths [13, 14], but some work has considered adaptive widths [15] or placement of the segment endpoints with genetic algorithms [16]. Methodologies of uniform segment sizes are not appropriate when nonuniform data duration distribution is present due in part to large fitting errors. Once the segments are created how to represent the data as values should be considered; methods such as aggregate [17], linear [7, 16], constant value [7], and symbolic [18, 19] are available.

Work with Piecewise Aggregate Approximation (PAA) started with the standard approach based on using the window size being fixed, as well as the width being determined from a single sample duration. The adaption of PAA to the data through the incorporation of genetic algorithms was done. This approach chose to focus on the fraction of the segments which contained values after conversion to PAA. This is different from the other literature approach with genetic algorithms and PAA that had a nonuniform sample distribution but a uniform number of samples when determining the boundaries [16], in addition the approach varied in the fitness function did not consider the representation outcome.

While some of the representation approaches have worked with nonuniformity, addressing data which has all multi-modal, nonuniform, time-limited time series properties has not been found thus far in the literature. The consideration of all the data properties is a core piece to the work that will be explored (either as the direct focus or as part of the data set used).

With every decision made in how the data is represented, the technique to determine a prognosis may be impacted. For example, in a Bayesian Network (BN), a larger domain of each input variable implies additional computation and a larger conditional probability table for each variable. Methods like Support Vector Machines (SVMs) utilize a mapping to feature space as part of the process. This implies how the data is represented and selection of the mapping function, the kernel, makes a difference

in the resulting classification.

An overarching idea recognized with this problem was to develop generalized quantifications for multi-modal, nonuniform, time-limited time series data. In exploring the literature, the quantification of data with Shannon's theorem being at the core [20, 21, 22]. Recent work has focused on different aspects of communication or the quantity of observations, such as the number of reviews for videos of an online site [23]. Slightly more quantity information is possible in other cases, such as Facebook being able to know the number of users and the number of friends (connections between users) [23]. The further extension of research done is tied to the storage of the large amounts of data physically [24]. Compressive and sparse sampling look at the quantity of data from the perspective of reconstruction necessities [25]. Comparisons have been observed at high level results, embedded in the optimization problems [16, 26, 27, 28], or through distances for similarity metrics [29, 30, 31].

In the chapters to follow relevant general background is first presented (Chapter 2) and work with further background will be presented when applicable in later chapters. The consideration of clinical samples extracted from the patient data and behavioral representation (Chapter 3) with classification and regression (Chapter 4) was then explored. Representing the patient data through PAA and evaluating with regression (Chapter 5) is then followed by the modification of the segmentation using genetic

algorithms based on the fit (Chapter 6). The recognition of the way to describe fit error with PAA and a segmentation approach are key novel contributions. Adaptations to work with multi-modal, nonuniform, time-limited time series data for classification and regression are also novel adaptations which have been made.

1.1 Contributions

Through the dissertation the following contributions were made:

† Development and application of discrete sampling approach combined with a simple behavioral representation approach on a multi-modal, nonuniform, time-limited time series data set evaluated with Bayesian Networks, Naïve Bayes, and Support Vector Machines. See Chapter 3. Presented in part at IEEE's Electro/Information Technology Conference 2016 and published in the proceedings [4].

† Development and application of discrete sampling approach combined with a simple behavioral representation approach on a multi-modal, nonuniform, time-limited time series data set evaluated with Linear and Quadratic Regression, Gaussian Process Regression and Support Vector Regression. See Chapter 4. Published in Advances in Science, Technology, and Engineering Systems Journal special edition issue on Recent Advances in Electrical and Electronics Engineering [32].

† Adaptation of Piecewise Aggregate Approximation to a multi-modal, nonuniform, time-limited time series data set and evaluated with regression. See Chapter 5.

† Development of a genetic algorithm approach based on sparseness feedback to improve the application of Piecewise Aggregate Approximation on a multimodal, nonuniform, time-limited time series with a focus on the performance in the evaluation of with regression analysis. See Chapter 6.

Chapter 2

Background and Relevant Work

For many applications the decisions made on how to represent data and the choice in evaluation technique impacts the performance objective. From Shannon, in classical sampling theory, through to the utilization of Support Vectors Machines, developed by Vapnik, these foundations contribute to the core background which is needed to pursue techniques to accommodate multi-modal, nonuniform, time-limited time series data. This section presents the background and relevant work on data representation and learning methods used and expanded upon in this work.

2.1 Data Representation

Data representation is critical to the utilization of observations and measurements that are collected in any application. In general, there are two main types of data, quantitative and qualitative, each of which have subtypes. Quantitative data, sometimes referred to as numeric data, deals with numbers or objective measures, e.g., height, weight, etc.; quantitative data is separated into discrete, or continuous data. Qualitative data includes data that comes from classifying or categorization; the three main types of qualitative data are binary, nominal, and ordinal data. There exist other alternative ways to describe types of data, e.g., in statistics the levels of measurement (nominal, ordinal, interval, and ratio) may be used. In addition to the type of data, another consideration is whether the data is aggregated. That is, is the data collected consisting of a single point, e.g., a patient's date of birth, or is it collected in a series, e.g., a patient's weight measured at each visit.

This work focuses on the application of outpatient medical data for analysis, where the raw observations and measurements can be described as binary observations, numerical observations, or time series. Binary data within the scenario of oncology, can be used with attributes such as life status (alive or dead) or a single event occurred (being born would be such an event).

Numeric value data is used to represent a quantity. Quantities can be observed either as a single measurement (such as, the age when death occurs) or an attribute which is time invariant (for example, the number of blood transfusions over a time). Numerical value data may be discrete, but can also be transformed to discrete. The numerical values across a population can be assigned to a limited set of discrete values using a rule based approach (values V_1 through V_2 become group $G1$ and V_2 through V_3 become group $G2$ for example) or through unsupervised learning techniques based on the data from a collection of samples.

The final observation type is time series. Background on the properties of time series data and representation approaches will be presented in the section to follow.

2.1.1 Time Series Data Representation

When observations are taken with respect to time, it is considered to be a time series. Time series values can form communication signals that are being transmitted, observations of a patients weight at different times, or the time of sunset each day.

A time series, $\mathbf{S}$, is a set of m discrete observations with each observation observed at a point in time, t, from some signal. The time series is defined by,

$$\mathbf{S}(t) = \{s_0, s_1, ..., s_{m-1}\}, \tag{2.1}$$

with time $t \in \mathbb{R}$ and $t = \{t_0, t_1, ..., t_{m-1}\}$. The observations are ordered with respect to time. The time between observations does not need to be uniform; an example of a nonuniform time series is seen in Fig. 2.1.

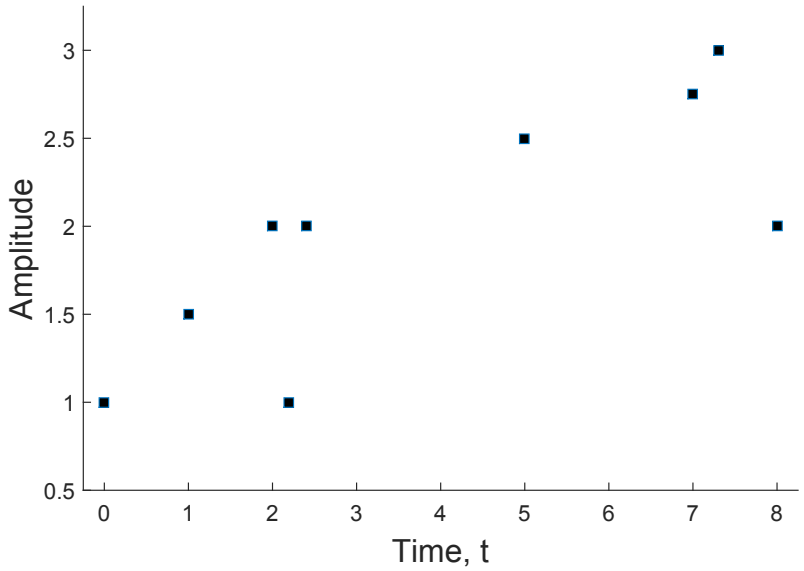


Figure 2.1: A nonuniform time series data sample. The x-axis represents time and the y-axis represents the amplitude of the observations.

Given a second time series, $\mathbf{R} = \{r_0, r_1, ..., r_{m_\mathbf{R}-1}\}$, the times of observations in $\mathbf{R}$ can be different from $\mathbf{S}$, like the two time series shown in Fig. 2.2. With the two time series, the nonuniform duration can be seen with the locations of endpoints of $[t_{n,first}, t_{n,last}]$, where n is the reference to the time series. From the difference between those endpoints the duration is calculated and equals $\Delta\, t_n$. In this example, the starting times for both $\mathbf{R}$ and $\mathbf{S}$ are at a similar time, but the ending time for $\mathbf{R}$ is earlier then $\mathbf{S}$. This duration relationship can then be described $\Delta\, t_\mathbf{S} > \Delta\, t_\mathbf{R}$.

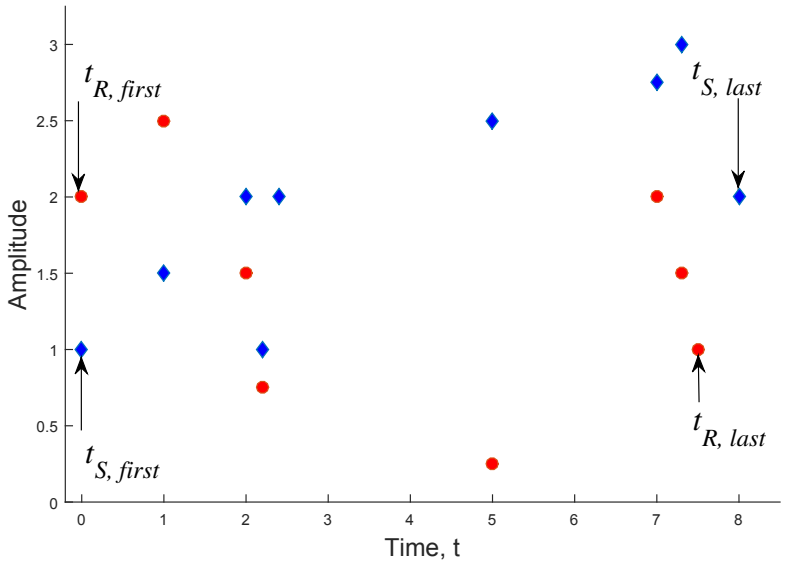


Figure 2.2: Two nonuniform time series with different durations as noted with the corresponding time points noted on the figure. Time series **S** is represented by the diamond marker and **R** by the circle marker.

Some examples of applications which have data that takes the form of a time series are fluid dynamics [33], electromagnetics [34], image reconstruction [35, 36], power systems [37], and motion and position monitoring [38].

Time series are comprised of observations, this makes the series able to be considered as a sample [39].

Samples can be multi-modal; for the time series sample, **S**, each observation represented would then have two indexing properties: the observation instance and the attribute series that the observation belongs to. Each attribute representing a time series is a dimension, resulting in k time series, and thus the number of k dimensions.

This modifies the notation to:

$$\mathbf{S} = \begin{bmatrix} s_{0,0} & s_{0,1} & s_{0,2} & \cdots & s_{0,m_0-1} \\ s_{1,0} & s_{1,1} & s_{1,1} & \cdots & s_{1,m_1-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{k-1,1} & s_{k-1,2} & s_{k-1,3} & \cdots & s_{k-1,m_k-1} \end{bmatrix}. \tag{2.2}$$

To allow for nonuniformity in the number of observations for each attribute, the number of observations, m, is independently defined for each attribute, k, creating the vector,

$$\mathbf{m} = \{m_0, m_1, ..., m_{k-1}\}. \tag{2.3}$$

Each of the k attributes also should be assumed to be independent of the other $k-1$ attributes in values and the times when the observations are made. In Fig. 2.3, a multi-model time series is presented with three attributes.

The observations for the time series modes thus far have been based on discrete observations made at a point in time, time series can also be used to represent other types of data. In image processing, the pixels and their respective color can be translated to a continuous time series [40]. For hand writing analysis, the height of the top of the characters form the time series [41]; making the time series essentially a tracing. In similarity recognition, the features of the shape can be extracted to form

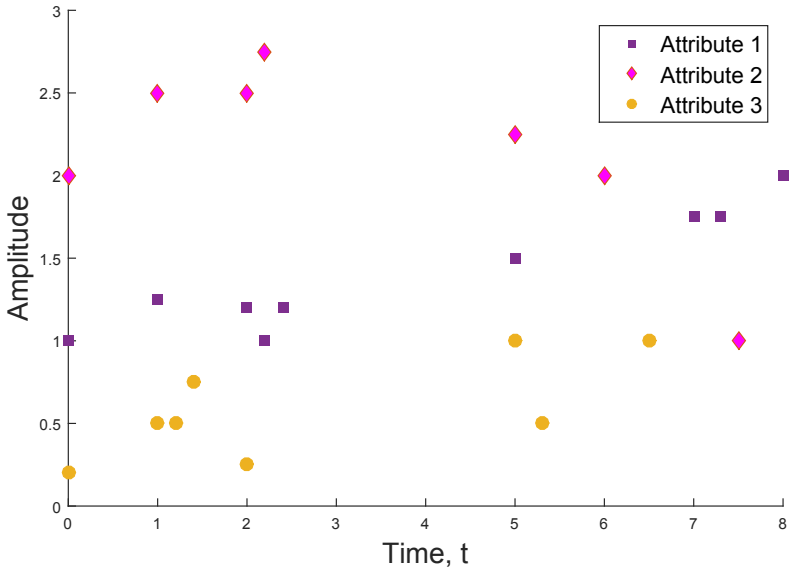


Figure 2.3: A multi-modal time series with three attributes.

a time series [42].

A large amount of research has been done on many approaches to represent time series, as displayed in Fig. 2.4 [1]. There is a division in representation approaches based on if the techniques are adaptive to the data or if they are a single rule format (for example, Discrete Fourier Transform (DFT) is presented as non-adaptive because the transform is a fixed formula applied to the data series with no variation based on the data). Formally, two definitions can be established.

Definition 1 *Data Adaptive representations: a common rule or approach for all data representation is chosen with the objective to minimizes the global reconstruction error [43].*

Definition 2 *Non-Data Adaptive representations: representation is based on an approximation from local properties of the data [43].*

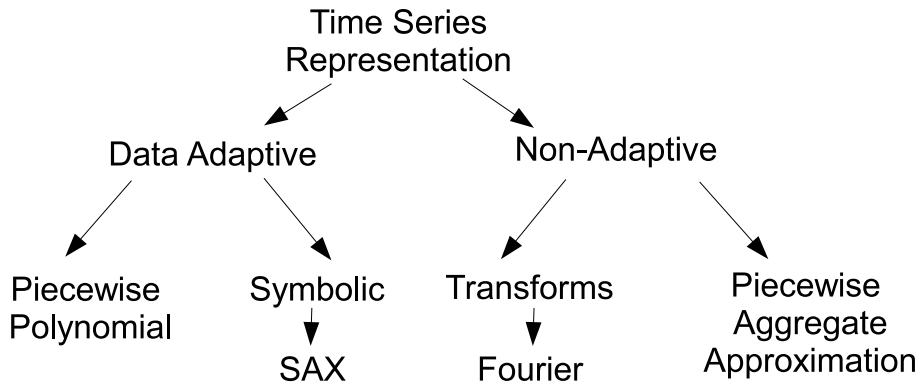


Figure 2.4: Taxonomy of different methods of representing time series data [1], see Appendix E.1 for copyright details.

Within any methods of data collection and data representation, missing data may be observed and methods to handle it must be implemented. For a sample with multiple time-series attributes, there is potential for a sample to lack the necessary observations. The missing observation can be all of the data for an attribute. If using a uniform sampling grid to describe the data, the sample can be missing an observation at one of the points which should be sampled.

Missing data and observations can take on the term of censored data when it is the result of the data for an event not being observed. In medical field data, there are three reasons for the data being censored: an event does not occur before the end of the study, an individual withdraws from further observations, or does not follow up (reasons include an individual's death [44]). Specifics on the outpatient oncology data

set to be used in the research are presented in Section 3.1.1, with the representation approaches discussed in Sections 3.1.2.1- 3.1.3.2, 5.3, 6.3, and 6.4.

2.1.2 Sampling Theory

When given a time series, the approach of what observations to consider for evaluation is often based on sampling theory. As previously defined a time series, **S**, is a collection of discrete observations made for an attribute that is continuous, making the original source of **S** known as the signal. The series **S** should consist of sufficient observations to describe the original signal. The determination of how often observations are needed and the methods regarding the relationship between the signal and series follow sampling theory. When data is uniform, selecting observations from the data can be done by following Shannon's Theorem [21]. In Shannon's Theorem, the frequency of observations to be able to represent the signal must be at minimum twice the frequency of the fastest feature. The frequency of the fastest feature is known as the Nyquist Frequency, this is used to properly select a sampling frequency Ω. From observations of **S** at the uniform intervals of $\frac{n\pi}{\Omega}$ (units are typically nanoseconds, microseconds, milliseconds, or seconds) samples can be takes to recreate the original signal $\hat{S}$.

With the spacing between samples described, the Whittaker-Kotel'nikov-Shannon

(WSK) sampling theory is formed. In the WSK Theorem, the ability to acquire uniformly spaced observations from **S** is needed. When there is no value obtainable at the desired time point additional considerations must be made. One approach to obtain uniformly spaced observations is interpolation. Using surrounding observations interpolation estimates the missing information needed to estimate an observation at that point of time. There are several interpolation approaches, often requiring a function to describe the signal to estimate the observation at the desired time point.

An alternative to interpolation is gridding. There are three steps in most gridding approaches: (i) the establishment of pre-weights, (ii) performing the convolution of the time series with the gridding kernel (which is often determined to have properties such as being shift-invariant and utilizes the Fourier Transform, FT) and (iii) then the inverse Fourier transform (IFT, background is presented in section 2.1.3) is performed to construct the new uniformly distributed set of observations [35]. The gridding kernel properties share the same principles as those used in Support Vector Machines (see section 2.3.3.1 for more background on kernels).

With the convolution's ability to resolve the non-continuous behavior of a sample, taking the integral or summation is a task which may still be very complex or difficult. One of the properties of convolution which simplifies the computation is the utilization of a transform. The Laplace Transform of a convolution results in the product of the Laplace Transform of each function individually [45]. This convolution property also

is applicable to the FT.

Using Shannon's Theory to determine what observations to consider, and WSK or gridding to reconstruct the original signal is a foundation in the pursuit of time series representation and in applications of kernels. Further explanation of WSK can be found in Appendix A.1.

2.1.3 Fourier Transformation

One of the tools used in sampling theory is the use of the Fourier Transform (FT). Additionally transforms are a representation approach for time series. In general, the FT takes a function or vector (in the discrete case) in the time domain and then transforms mathematically to the frequency domain through integration.

For the discrete form, the integral is replaced with a summation. The FT is then known as the Discrete Time Fourier Transform (DTFT) [46]. Moving from the time domain to the frequency domain has its benefits, primarily the inclusion of a periodic function that makes a sample that can be uniformly sampled possible (gridding). Once the function is periodic, the return of the sample to the time domain is needed to allow for the results to be used, this is done using the Inverse Fourier Transform (IFT) or Inverse Discrete Fourier Transform (IDFT).

DFT is used to reduce dimensionality by considering only a subset of coefficients and then normalized allows for invariance to rotation and starting points [47]. Methods of representation of time series based on FT have been applied in applications such as fluid dynamics [33], electromagnetics [34], image reconstruction [35, 36]. In image recognition, such as finding similar images from a query in a database, filtering techniques can be used to account for variations which don't factor into the decisions, such as the orientation of a shape [42, 47].

An extension of the computation of the DFT is the Fast Fourier Transform (FFT) which performs the transform quicker then a traditional FT. Outside of the traditional FT, FFT, and DFT, there are a series of nonuniform Fourier Transform (NUFFT, as they most commonly utilize the FFT as the basis) which exist. There are three primary types: type 1 one is when the transform is used for an original system which is irregular to achieve a regular grid, type two takes a uniformly sampled system and produced an irregular grid, and then type three takes one irregular grid and turns it into another irregular grid [34].

In addition to FTs, there are other transform methods which can be used. Other transforms include the z-transform [48], discrete cosine transform [48], the FFT [48], discrete Hilbert Transform [48], Walsh Transform [37], and wavelet transform [49].

In many applications the utilization of the transform lies with the coefficient not with the signal recreation. The lower or initial coefficients are found and maintained to be

used in the evaluation and higher frequency coefficients are disregarded [50].

Additional detail details on Fourier Transformation are included in Appendix A.1.1.

2.1.4 Piecewise Aggregate Approximation

Segmentation representation is used to simplify the data available by dimensionality reduction or compression sampling. In the methods based on segmentation, the time series is broken down to c segments. The regions within the segments are then represented through some approach, such as linear, cubic or aggregate methods [7]. The aggregate technique is known as Piecewise Aggregate Approximation (PAA) [17].

The segment window (also referred to as window, segment, frame or piece) represents a slice of the information of some duration (or width). In general, a window is summarized with a vector of features describing the behavior of that region [51].

For each of the c windows for a time series, n, l_n is the window width, that can be defined relative to the duration of the sample,

$$l_n = \frac{\Delta\, t_n}{c}. \tag{2.4}$$

The placement of the segments are determined by split points which are traditionally

places with equal spacing across the duration $\Delta\ t_n$.

Nonuniform data presents PAA with the situation that a window or many windows, even adjacent, may not contain any observations of the signal. To mediate this complication, one approach is to utilize the neighboring windows that have observations and use the PAA values from those windows as the two values to compute the average. The computed average will become the value of the window in place of missing observations. This interpolation approach is not going to create an accurate behavioral representation depending on the distribution of observations and the window size.

In one recent work, the determination of the segment boundary placements has been considered with genetic algorithms for samples that are discrete and equal count of observations through the entire population to then be applied to the entire population [16]. Working with electrocardiogram data, a evolutionary computation approach was used to determine the placement of segment boundaries through an optimization problem. This is an University of California Riverside (UCR) data set, which is a classification time series time set that is a single mode time series of uniform length [52].

2.1.4.1 Symbolic Aggregate Approximation

Symbolic aggregate approximation (SAX) [19] is an extension of PAA, when the value based assignments in the segments are transformed to symbolic assignments.

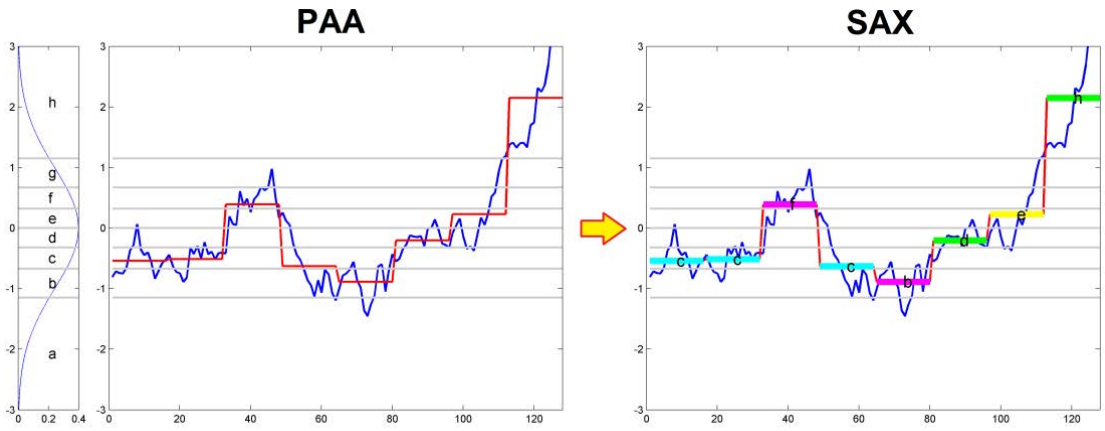


Figure 2.5: A visual representation of break-point assignment to transform PAA to SAX [2]. © 2012 P. Montalto, M. Aliotta, A. Cannata, C. Cassisi, and A. Pulvirenti. Adapted from "Advances in Data Mining Knowledge Discovery and Applications", under CC by 3.0 license. Available from: doi.org/10.5772/49941.

Symbolic assignment approaches utilize symbols (such as an alphabet) in place of values.

There are three primary steps to implementing SAX. The normalization of the values to have a zero mean and standard deviation of one is performed to prepare the data. Once the representation has been shifted in preparation then the creation of the windows is conducted. These windows are of length l_n and should be uniform in length. With these windows the normalized information is then represented with PAA. The final step of SAX is the conversion of the discretization of the window's by mapping to discrete symbols [53].

For a time series $\mathbf{S}$ of m observations (making the series length equal m), with a set of v discrete values to be assigned the SAX of $\mathbf{S}$ can be determined over an aggregate

length segment, l_{agg}. The term l_{agg} corresponds to the aggregate value for the segment l, that combining all l makes a new series. This first step is of determining l_{agg} is the application of Piecewise Aggregate Approximation. Once the regions are represented in the aggregate form, symbolic assignments are made from an alphabet set. The breakpoints assigned provide regions of values that correspond to each letter in the alphabet set used for the symbolic representation assignments.

Fig. 2.5 summarizes the process with the Gaussian distribution on the left covering the entire set of values that is used to take the PAA on the middle plot to assign the symbolic alphabet. This results in the SAX plot on the right.

As the segments are represented symbolically, definitions of distance between symbols need to be defined to determine similarity when using distance calculations between time series. A look-up table developed to use distance definitions when calculating distances [18].

With SAX, the representation is appropriate when there are not substantial changes in amplitude within a segment. When large amplitude variations occur, the aggregate value determined may not adequately represent the behavior of the segment, with extreme values potentially shifting the symbolic assignment. The extreme points may also be lost in the process. One work proposed the extension of SAX as a triple tuple, that the maximum, minimum, and mean are considered in the representation. This provides more information about the behavior in the representation [14]. The cost of

the additional information is the increases the dimensionality needed for representation.

With SAX there have been many expansions considered to improve the performance. Expansions have predominately focused on the application or the look up table component or break point boundaries for symbolic assignment [14, 54]. These expansions in some cases utilize other methods including genetic algorithms. A genetic algorithm can be used to create a look-up table for break point placement hat will be used for the the discretization step to assign symbols [18].

2.1.5 Distances

In data representation (and many machine learning or data mining) techniques, there is often a similarity consideration that is done using a distance calculation. To calculate the distance between two points (or points on lines), Euclidean Distance is often used. For two points, x_1 and x_2, then the Euclidean Distance, $d(x_1, x_2)$ is calculated by [55],

$$d\left(x_1, x_2\right) = \sqrt{x_1^2 + x_2^2}. \tag{2.5}$$

This can then be generalized to the Minkowski Distance, or as the L^p norm [51],

$$L^p(x_r, x_s) = \left(\sum_k |x_{r,k} - x_{s,k}|^p \right)^{1/p}, \tag{2.6}$$

where k is a the number of elements being compared. For a pair of points, $k = 1$. For a pair of signals, k is the number of elements in the set of vectors describing the signals.

With the generalized Minkowski Distance equation, the Euclidean Distance is the case where $p = 2$. Another common distance utilized is Manhattan Distance [51], which is when $p = 1$. The maximum distance is defined with L_∞ [56]. Distance calculations are often used in determination of similarities between two points or signals [31, 57]. Additional similarity and distance metrics include Jaccard Distance [58], Hamming Distance [59], cosine angle distance [60], and Mahalanobis Distance [55, 61].

When there is a time shift or scaling variance between the time series being compared, similarities can be obscured. To improve the recognition of similarity in these cases, a method developed in the 1970's for speech recognition known as dynamic time warping (DTW) can be used [56, 62]. Details on DTW is explained in Appendix A.2.

2.1.6 Temporal Abstraction

An alternative dimensionality reduction technique that is applicable to time series is temporal abstraction. Temporal abstractions can be based on value based or trend based information [63], producing a high level qualitative representation [64]. Overall the idea is to reduce the information available down to behaviors over intervals and

use inference to supplement the missing information. One of the key features to make the work possible is the description of the data with a zero-point time stamp, which may in reality may not be a real world value of zero. Meaning the zero point may be an age or time [65].

For each variable or mode there are two parts to the representation; the interval and the value. The interval is the start and end time points for the region that will be assigned values, for example $[b_1, e_1]$ being the first interval. To create the segmentation intervals, one method is sliding windows. Sliding windows starts with smaller windows intervals and increases the width until an error is reached [64]. Time intervals can also be predefined and then merged when the abstraction of adjacent regions are the same [65]. In predefined interval systems it is possible for regions without observations to occur, in those cases interpolation or use of a persistence function based on surrounding information can be used [65].

The value takes on an assignment from an alphabet relative to what is being represented. In the case of trend abstraction decreasing (D), steady (S), and increasing (I) are used [64]. Similarly for value based abstraction using percentile based lab values can be used to assign values such as high (H) or low (L), with as many classes as desired [64]. Overall the state, gradient, and rate can be used to describe the abstraction value [65]. For each variable the set of observations becomes simplified to $< v_1[b_1, e_1], ..., v_n[b_n, e_n] >$, when there are n intervals [64].

The most common usage of temporal abstraction is through combining with Allen's Temporal Relations [65]. This is how to describe relationships between the intervals; such as before, during or after. For example a case may be that the patient exhibits increasing hemoglobin while the weight is stable in an temporal abstraction with Allen's temporal relations used.

Pattern matching is a common application which also makes use of temporal abstraction. This would then take the pattern of interval behaviors and relationships to match their patterns with known patterns symbolic of outcomes [64]. Patterns play a large role with this representation approach, with the consideration of mining minimal predictive temporal patterns, this is analogous to feature selection [63] and necessary to avoid undesirable classification results.

2.2 Common Data Sets

One of the things that sets the work being approached apart from much of the literature is the use of multi-modal, nonuniform, time-limited time series data. In many of the referenced papers, the data sets come from the University of California repositories hosted at University of California Riverside (UCR) [52] or University of California Irvine (UCI) [66].

The UCR data sets are all time series sets, however they are all uniform in length and observed at regular observational rates with no missing observations. In addition the UCR set objective is classification. One other difference with these sets is that preprocessing or z-normalization has already been done [52]. One last difference to emphasize is that these sets are all single mode. As on August 2016, there were 85 different types of single mode time series data sets within this archive.

The UCI repository has a much more vast set of information. In the closest sets to the work being done, the sampling is performed at a known frequency making traditional interpolation a practical solution. There is no set which was found in those repositories that approaches the degree of complexity of the data set being used in this dissertation.

2.3 Classification

One of the fundamental areas of machine learning is the process of classification [67], the objective of predicting a qualitative output [68]. Some examples where classification is applicable in the oncology domains includes the determination of prognosis, forecasting, and population descriptions (such as demographic highlights of regions). In this section, some machine learning techniques which are applicable to classification will be presented.

In general, the classification problem is defined as given a sample and a set of labels (two or more options), assign a label to the sample. For example, a sample, $\mathbf{x}$, contains n values. The class label of y is one of the options such that $y \in \mathbf{Y}$. The classification problem aims to use the values of $\mathbf{x}$ to select y from $\mathbf{Y}$ using a classification technique.

Binary classification is where the outcome of classification being one of only two options in $\mathbf{Y}$. Multi-class classification is when there are more then two possible class labels which can be selected as the classified outcome, more then two options in $\mathbf{Y}$. Some methods such as NB are capable of multi-class classification other methods require combinations of binary classifications.

Two approaches for multi-class classification using binary classifiers are One-versus-One or One-versus-All, in both the class with the highest accuracy is the label selected.

One-versus-One approach performs binary classification between two classes at a time and looks at all the pairwise combinations. When there are k classes, then there will be a total of $k\left(k-1\right)/2$ classifier evaluations conducted. In One-versus-All approach to classification, each class is represented as an individual classifier against the union of the other classes in a binary classifier approach. This is done for each class, so for k classes there will be k classifier evaluations run [69, 70].

Some variations of supervised learning classification methods studied in this work include Naïve Bayes, Bayesian Networks, and Support Vector Machines.

2.3.1 Naïve Bayes

One model which can be used in supervised learning for classification is Naïve Bayes (NB). A NB model can be represented as a Bayesian Network, where each attribute is incorporated in the network as a node. In this model, it is assumed that all attributes are conditionally independent of each other given the class label [51]. The NB network is configured as the example in Fig. 2.6 with a central node connected to all the other nodes. The central node is the node of interest, representing what is class variable, Y. The result of the NB is the assignment of a classification label to Y.

Consider a classification data set, $\mathcal{D} = \{(\mathbf{x}_k, y_k)\}, k = 1, \ldots, n$, consisting of input samples, $\mathbf{x} = (x_1, x_2, \ldots, x_m)$, drawn from the variables, $X_1, X_2, \ldots, X_m$, and a class

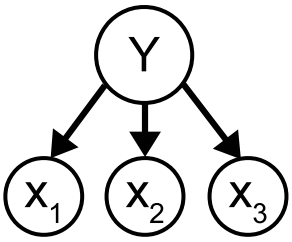


Figure 2.6: Example of a NB network, this one is an example of the network used in the preliminary work with the oncology prognosis data set.

label, $y_k \in \mathbf{Y}$. The Bayesian classifier looks to compute the posterior probability, $P(y_k \mid X_1, X_2, \ldots, X_m)$, for all k values of y and select the class that maximizes this probability. Using Bayes Rules,

$$P(y_k \mid X_1, X_2, \ldots, X_m) = \frac{P(X_1, X_2, \ldots, X_m \mid y_k)P(y_k)}{P(X_1, X_2, \ldots, X_m)}. \tag{2.7}$$

The NB classifier makes use of the strong assumption of conditional independence among the variables given the class, resulting in the following modification of the calculation of the posterior probability,

$$P(y_k \mid X_1, X_2, \ldots, X_m) = \frac{\prod_i P(X_i \mid y_k)P(y_k)}{P(X_1, X_2, \ldots, X_m)}. \tag{2.8}$$

The denominator of the fraction is that same for all classes, y_k, therefore the probability is proportional total,

$$P(y_k \mid X_1, X_2, \ldots, X_m) \propto \prod_i P(X_i \mid y_k)P(y_k). \tag{2.9}$$

The training of the NB model involves estimating the probabilities above from the data set. New data samples can be evaluated using the learned likelihood probabilities. The class is determined as the one with the highest probability resulting from the evaluation. This calculation can be used to select the best class from $\mathbf{Y}$, where $|\mathbf{Y}| = d$, calculate,

$$y = \underset{y_i \in \{y_0 \cdots y_{d-1}\}}{\arg\max} P(Y = y_i \mid \mathbf{X} = \mathbf{x}). \tag{2.10}$$

2.3.2 Bayesian Networks

Bayesian Networks (BNs) are one method that is used for classification, relying on inference to determine the assignment of class. In a BN, the network is represented through a directed acyclic graph (DAG). The DAG represents the properties of conditional independence between variables [71, 72].

Each node is representing an attribute or variable, that is connected to one or more other nodes if it has any influence on the probability of that node. The edges on the DAG connect the nodes as a graphical representation of the influences within the network. For all BN, the Markov condition holds that is, for each node/variable it is conditionally independent of all it's non-decedents given the set of it's parents [73]. The result of the BN is a graphical representation of probabilistic relationships which

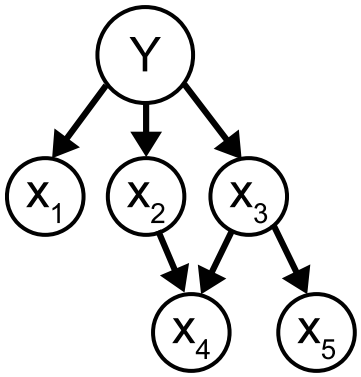


Figure 2.7: A DAG to visualize a BN.

can be used for inference [73]. Each node has a related conditional probability function, which describes the probability of each of the values in the domain for the random variable. The classification of the variable is then made by the Maximum a Posteriori (MAP) as done with NB, only additional considerations must be made to account for the additional nodes connected in the network.

The use of BN requires algorithms or domain expertise to learn the network. The problem of learning the structure of a BN has the designation of NP Hard [71].

An example of a DAG representing a BN is seen in Fig. 2.7. Once the network structure has been determined then the conditional probability tables for each node of the network can be estimated from data or domain knowledge.

After the learning occurs and the BN has been fully specified (network structure and conditional probability tables), then new samples may be evaluated to determine the label. The label being selected using the same approach as with NB, selecting the

label with the highest probability (MAP).

There are many different tools available for the learning and evaluation of BN [74]. In my research, a graphical based tool, GeNIe was used that was developed by the University of Pittsburgh. GeNIe (Graphical Network Interface) is software package [75, 76] that learns the parameters of the network through Expectation Maximization (EM) within the built-in functions. The test data is evaluated on the trained network in GeNIe, where the length of survival category with the maximum probability is selected, then analyzed with the mean and standard deviation across the folds.

2.3.3 Support Vector Machines

Pioneered by Vapnik, support vectors have been used for both pattern classification and regression. Support Vector Regression (SVR) is the approach applicable for function estimation (explained further in Section 2.4.1). Support Vector Machines (SVM) is the method utilized with classification [77].

The SVM process begins with a sample of data, $\mathbf{S} = \{(\mathbf{x}_k, y_k)\}, k = 1, \ldots, l$, consisting of input samples $\mathbf{x} = (x_1, x_2, \ldots, x_m)$, and a class label $y \in \{-1, +1\}$. The goal of the SVM is to find a linear hyperplane to separate to two classes of samples, where the hyperplane is defined as,

$$f(\mathbf{x}) = sign(\mathbf{w} \cdot \mathbf{x} + b). \tag{2.11}$$

The optimal separating hyperplane is selected to maximize the margin ρ between the two classes of data. In Fig. 2.8, a separating hyperplane is showed with the margin that it creates between the two classes. The support vectors are data samples (vectors) which are located on the parallel hyperplanes used to calculate the margin distance. Specifically, a support vector, $\mathbf{x}_i$ is when in $\mathbf{x}$ at instance i, $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) = 1$, with y_i being the label for instance i [78].

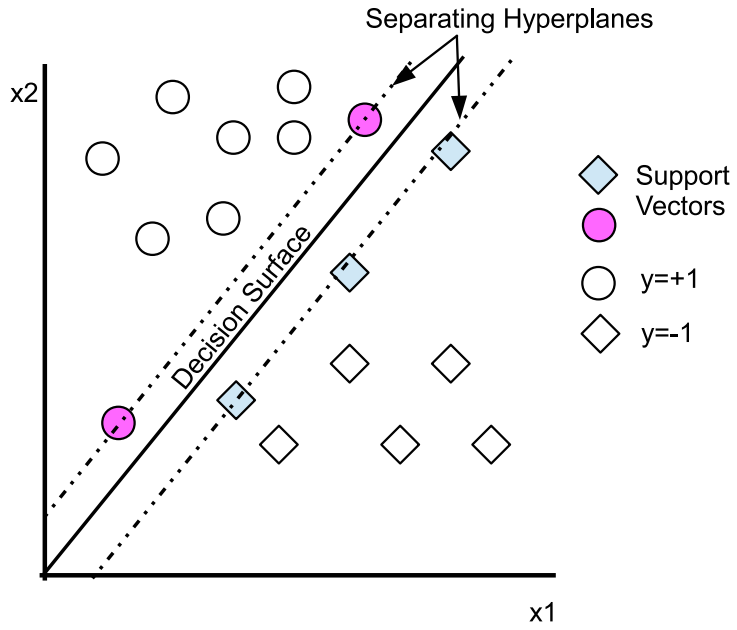


Figure 2.8: Two classes (y=+1 and y=-1) which an SVM hyperplane is placed in a manor that is linearly separable. The support vectors are the points which are colored and are used to determine the placement of the hyperplane [3].

Many extensions are available that take this base concept or the SVM problem and develop it in two different aspects: space transformation and optimization modification.

Space transformation is an approach that can be known as the kernel trick [51], transforming the space which the evaluations are considered from input space to a higher dimension feature space. The transformation is done using kernels, $K(x, x_i)$, where x is the input vector [78, 79]. Use of the kernel trick allows for non-linearly separable data in input space, to be separated by the linear hyperplane in the feature space [78]. When the mapping produces a complete feature space, that space is known as a Hilbert Space [80].

The optimization modification is made through the incorporation of additional parameters. The slack variable, ξ, is used to modify the approach to a soft margin classifier [78]. The role ξ plays in the optimization is to provide misclassification penalty component, in conjunction with an explicit cost parameter, C.

The optimization problem accounting for the slack variable, cost variable, and the kernel trick is now developed with the bounding constraints for the sample instances $(\mathbf{x}_i, y_i)$ with $i \in \{1, ..., l\}$, and y_i is the label with values $\{-1, 1\}$. The slack and cost variables must both be greater then zero [78, 81, 82],

$$\begin{aligned} &\min_{\mathbf{w},b,\xi} \ \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{l}\xi_i, \\ &\text{s.t. } y_i(\mathbf{w}^T\phi(\mathbf{x}_i) + b) \geq 1 - \xi_i. \end{aligned} \tag{2.12}$$

In this equation, ϕ is the function that allows the mapping to a higher dimensional space.

Alternative approaches to view the optimization problem lend itself to the ability to consider different parameters and even make the problem easier to solve [83]. One method is that of the dual problem formulation is done through Wolfe Principle and

Karush-Kuhn Tucker principle [82, 83, 84, 85],

$$\begin{aligned} h(x) &= sgn(< \mathbf{w} \cdot \mathbf{x} > +b) \\ &= sgn(\sum_{j=1}^{l} \alpha_j y_j < \mathbf{x}_j \cdot \mathbf{x} > +b) \\ &= sgn(\sum_{j=1}^{l} \alpha_j y_j < \mathbf{x}_j \cdot \mathbf{x} > +b). \end{aligned} \tag{2.13}$$

In this dual representation the α parameter (the Lagrange Multipliers) is not zero only for the support vectors. This creates a sparse vector when evaluating the SVM.

Additionally variations on SVM can be made, in a semi-supervised learning environment the ε-loss function is used as an influencing parameter to the optimization to bound the errors in the margin developed [86]. There is the ν-SVM which incorporates a ν parameter to assist in the tuning of the parameters [85].

Similar to the SVM classification problem is support vector regression, where value outcomes are determined. Forecasting of time series is very similar to classification of an expected future outcome and have used SVMs for evaluation methods [87, 88]. For multi-class classification, SVMs often use either the One-versus-One or One-versus-All approaches.

2.3.3.1 Kernel Functions

The kernel function takes many forms and is used in many applications (such as gridding). Kernels are used in SVMs to evaluate the data in a higher dimension without performing explicit mapping. The kernel function is a covariance [89], that is defined by [83, 87],

$$K(\mathbf{x}, \mathbf{z}) = \langle \phi(\mathbf{x}), \phi(\mathbf{z}) \rangle. \tag{2.14}$$

For kernels to be of a valid form, it is desired that the Mercer conditions are satisfied. Mercer's theorem is formally defined as:

Theorem 1 *To guarantee that a continuously symmetric function* $K(\mathbf{u}, \mathbf{v})$ *in* L_2 *has an expansion:* $K(\mathbf{u}, \mathbf{v}) = \sum_{k=1}^{\infty} a_k z_k(u) z_k(v)$ *with positive coefficients* $a_k > 0$ *(i.e.* $K(\mathbf{u}, \mathbf{v})$ *describes an inner product in some feature space), it is necessary and sufficient that the condition* $\int_C \int_C K(\mathbf{u}, \mathbf{v}) g(u) g(v) du dv \geq 0$ *be valid for all* $g \in L_2(C)$ *where* C *is a compact subset of* R^n*. The expanse property means that for the function* $K(\mathbf{u}, \mathbf{v})$ *the right hand side converges absolutely and uniformly [82].*

To confirm the kernel, $K(\mathbf{x}, \mathbf{z})$, satisfies the Mercer condition there are two properties we look at. First is the commutative property, which the order the mapping of the

vectors $\mathbf{x}$ and $\mathbf{z}$ are when the inner product are invariant. The expression,

$$\langle \phi(\mathbf{x}) \cdot \phi(\mathbf{z}) \rangle = \langle \phi(\mathbf{z}) \cdot \phi(\mathbf{x}) \rangle, \tag{2.15}$$

demonstrates this property [87].

The second condition to the Mercer conditions is that the kernel needs to be a positive semi-definite (PSD) type [83]. A PSD association is given to a system in which all the eigenvalues are non-negative. This implies $\forall \alpha, \alpha^T Q \alpha \geq 0$.

The positive consideration can be expressed as $\int_{X \times X} K(\mathbf{x}, \mathbf{z}) f(x) f(z) dx dz \geq 0 \forall f \in L_2(X)$, that the subsets must also be positive [83].

Additionally we look to the satisfaction of Cauchy-Schwarz inequality,

$$K(\mathbf{x}, \mathbf{z}) = \langle \phi(\mathbf{x}) \cdot \phi(\mathbf{z}) \rangle^2 \leq \| \phi(\mathbf{x}) \|^2 \| \phi(\mathbf{z}) \|^2 . \tag{2.16}$$

The Cauchy-Schwarz inequality is then simplified to the condition,

$$K(\mathbf{x}, \mathbf{z}) = \langle \phi(\mathbf{x}) \cdot \phi(\mathbf{x}) \rangle \langle \phi(\mathbf{z}) \cdot \phi(\mathbf{z}) \rangle = K(\mathbf{x}, \mathbf{x}) K(\mathbf{z}, \mathbf{z}) . \tag{2.17}$$

The satisfaction of Cauchy-Schwarz inequality is analogous to the satisfaction of the

triangle inequality. This relationship provides a bounding to the results along with a set of substitutions available.

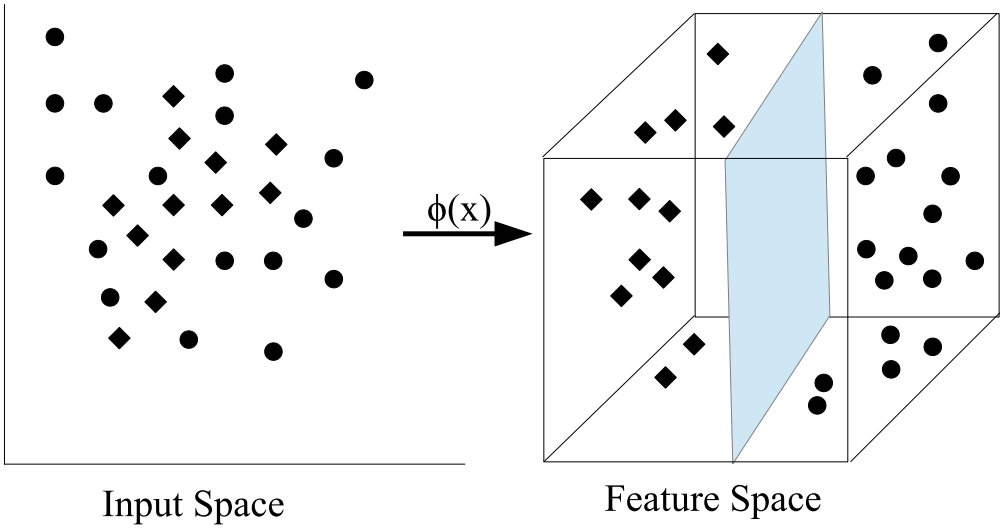


Figure 2.9: The mapping from input space to feature space, allowing for a hyperplane to linearly separate the circles and diamonds when separation was not initially possible in input space.

Following the semi-definite positive matrix and Mercer's rules, there are many commonly used kernel functions for SVMs. Some of these kernels are presented in Table 2.1.

Table 2.1
COMMON KERNELS USED WITH SVM [5].

Kernel Name	**Kernel**
Linear	$K(\mathbf{x},\mathbf{y}) = \mathbf{x}^T\mathbf{y} + c$
Polynomials: For some positive integer d	$K(\mathbf{x},\mathbf{y}) = (1 + \langle \mathbf{x},\mathbf{y}\rangle)^d$
RBF/Gaussian	$K(\mathbf{x},\mathbf{y}) = \exp\left(\frac{\langle(\mathbf{xy}),(\mathbf{xy})\rangle}{(2\sigma^2)}\right)$
Sigmoid	$K(\mathbf{x},\mathbf{y}) = \tanh\left(\alpha\mathbf{x}^T\mathbf{y} + c\right)$
Cauchy	$K(\mathbf{x},\mathbf{y}) = \frac{1}{1+\frac{\|\|\mathbf{x}-\mathbf{y}\|\|^2}{\sigma^2}}$
Wave	$K(\mathbf{x},\mathbf{y}) = \frac{\theta}{\|\mathbf{x}-\mathbf{y}\|} sin\frac{\|\mathbf{x}-\mathbf{y}\|}{\theta}$

With kernels that follow Mercer's rules, if combined the resulting kernel is still a Mercer Kernel [90]. Some approaches of developing a new kernel based on combining two known Mercer Kernels (the positive and semi-definite conditions are met) [83]. For example, kernels $K1$ and $K2$ can be used to develop a new kernel $K3$ in two ways:

$$\begin{aligned} K3(\mathbf{x}, \mathbf{z}) &= K1(\mathbf{x}, \mathbf{z}) + K2(\mathbf{x}, \mathbf{z}) \\ K3(\mathbf{x}, \mathbf{z}) &= K1(\mathbf{x}, \mathbf{z}) K2(\mathbf{x}, \mathbf{z}) \end{aligned} \tag{2.18}$$

One further alternative to combine kernels beyond arithmetic is to use multiple kernels. A single kernel can replaced with a mixture of kernels [91]. In this approach the inputs $\mathbf{x}$ and $\mathbf{z}$ of $K(\mathbf{x}, \mathbf{z})$ are divided into subsets to correspond to each of the kernels. Each of the kernels are weighted, η_j, $j = 0, ..., m-1$ with m kernels, and has the constraint of $\sum_{j=0}^{m-1} \eta_j$.

2.4 Regression

The problem of regression is a supervised learning technique that aims to develop a model to map an input $\mathbf{x}$ to an output $f(\mathbf{x})$, predicting a quantitative output [68]. The assigned output is a prediction of a continuous quantity or numerical value. Optimization of the fit of the model is done through the minimization of error. Depending on the complexity of the model, techniques like simulated annealing and genetic algorithms may be used for the optimization process or simple analytical analysis [91].

One example of regression is a linear regression model which takes the form,

$$f(\mathbf{x}) = \mathbf{w}\mathbf{x} + w_0, \tag{2.19}$$

where $\mathbf{x}$ is the input and $\mathbf{w}$ is the weight that fits the model, that for a linear model is the slope. The parameter w_0 is the offset or bias parameter to adjust the fit. The parameters in this case are chosen based on the minimization of the error when fitting with the training set. One way that the parameters can be selected is through the maximum likelihood estimation [91].

Other polynomial functions are very common techniques for regression, including higher order. For example, the quadratic regression approach is one order higher

then the linear regression approach. It takes the form as follows,

$$f(\mathbf{x}) = \mathbf{w}\mathbf{x}^2 + \mathbf{z}\mathbf{x} + \mathbf{b}. \tag{2.20}$$

The process of regression can be modified to select coefficients to improve model performance in an approach known as stepwise regression. Through the process of determining the coefficients are added and removed from the model with the intention to improve the fit [92]. The approach trades off forward selection and backward elimination, until all possible variables are attempted to be added [93]. Stopping criterion, such as partial sum of squares or t-tests, can be used to stop the process of adding or deleting more variables. [93, 94]

Regression can be implemented with many approaches beyond polynomials; including kernel machines, support vector machines, and Gaussian Processes.

2.4.1 Support Vector Regression

Support vector regression (SVR) is a kernel based regression approach. The regression is a set of linear functions,

$$f(\mathbf{x}, \alpha) = (\mathbf{w}, \mathbf{x}) + \mathbf{b}, \tag{2.21}$$

that is aimed at minimizing the loss function ε, and where α is the Lagrange multiplier. The support vectors are represented through the term $\mathbf{x}$, and during the fit process variables $\mathbf{w}$ and $\mathbf{b}$ are determined, such that $\mathbf{w}$ corresponds to the weight and $\mathbf{b}$ is the offset or bias. To allow for the spread in values, a slack variable, ξ_i is used. The objective is then to minimize the function [79],

$$\Phi(\mathbf{w}, \xi^*, \xi) = \frac{1}{2}(\mathbf{w} \cdot \mathbf{w}) + C\Big(\sum_{i=1}^{l} \xi_i^* + \sum_{i=1}^{l} \xi_i\Big), \tag{2.22}$$

when there are l samples. To support this boundary, the slack variable ξ_i must be greater then or equal to zero. In the evaluation, the constraint is used to relate the loss and slack variables to the function [79],

$$y_i - (\mathbf{w} \cdot x_i) - \mathbf{b} \leq \varepsilon + \xi_i, i = 1, ..., l. \tag{2.23}$$

The SVR approach can be extended to allow for the application of kernels which satisfy Mercer's condition to be used [79]. Some examples of kernels include linear and radial basis functions.

2.4.2 Gaussian Processes

Just with standard regression approaches, the aim is for the model to be learned from a set of inputs with known outcomes. With a Gaussian Process (GP) the inputs are

treated as a set of random variables (RV) are incorporated with a covariance function to determine a probabilistic outcome and the regression value [95]. A RV is a variable whose value is measurable [84].

The GP model is defined through the mean function and covariance function, based on the collection of RVs that have a joint probability distribution that is both consistent and Gaussian. In the model, a kernel based covariance function describes the fit behavior. This approach is based on Bayesian theory, the posterior is computed to determine the regression outcome [96].

For a set of training data with n samples, $\mathcal{D} = \{(\mathbf{x}_k, y_k)\}, k = 1, \ldots, n$, in which y_k is a value and $\mathbf{x}_k$ drawn from the variables $X_1, X_2, \ldots, X_m$, are also known as the RVs in the GP. The outcomes are then summarized to, $Y = (y(\mathbf{x}_1), y(\mathbf{x}_2), ...)$ with a Gaussian distribution with the probability distribution calculated by [97],

$$\mathbf{P}(Y|C, \{\mathbf{x}_i\}) = \frac{1}{Z} exp(-\frac{1}{2}(Y - \mu)^T C^{-1}(Y - \mu)), \tag{2.24}$$

with μ being the mean vector and C being the covariance matrix. The variable Z is a latent/hidden variable. By assuming a zero mean, hence $\mu = 0$, this distribution can be used to predict the outcome, y of an input $\mathbf{x}$ with the training set $\mathcal{D}$,

$$\mathbf{P}(y|\mathcal{D}, \mathbf{x}) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(y - y*)^2}{2\sigma^2}). \tag{2.25}$$

Such that,

$$y* = k(\mathbf{x})^T C_n^{-1}(y_1, ..., y_n), \tag{2.26}$$

is formed that C_n is the covariance matrix trained from the n training samples and,

$$k(\mathbf{x}) = (C(x_1, \mathbf{x}), ..., C(x_n, \mathbf{x})), \tag{2.27}$$

is the covariance matrix between the sample inputs and the input being evaluated to determine the regression result. The variance term, σ, incorporates the covariance function applied to the input,

$$\sigma = C(\mathbf{x}, \mathbf{x}) - k(\mathbf{x})^T C_n^{-1} k(\mathbf{x}). \tag{2.28}$$

The covariance function can be switched with a kernel function to add smoothing, periodicity, or generality to the behavior [97]. This method in general allows for the merger of the kernel trick with Bayes rule to compute the regression outcome [95].

2.5 Genetic Algorithms

Genetic Algorithms (GAs) are a common approach to optimization based on the biological idea of genetic evolution. The approach owes its roots to the work done by Fraser, Bremermann, Reed, and Holland [98]. The more popular canonical GA approach used proposed by Holland [98]. The driving operators are selection (aka survival of the fittest) and recombination [98].

The functional parameters within GAs include the fitness function, population parameters, and stopping conditions. There can also be bounding constraints. In the GA, the representation approach in the chromosomes is also a key parameter, some approaches limiting genetic operator functions that can be used in creating new generations.

The fitness function is the optimization function used in the selection of the best chromosome. It can be used via minimization [99] or maximization [100].

The population parameters describe how the population of chromosomes exists and changes from generation to generation. The chromosome can be represented in many ways. In general the structure is a string of length l, making it an l-tuple [101]. The l terms are known as genes and each gene is taken from a set of values known as alleles [102]. The values of the alleles can be encoded in binary, floating point or other

approaches. In any representation the chromosomes represent the characteristics of the individual [98].

The parameters for the GA implementation include the maximum size of the population and genetic operators. Genetic operators include mutation, crossover, inversion, dominance modifications, translation, and deletion [101]. Mutation rate describes the random changes that would occur. An example of mutation is seen in Fig. 2.10. Mutation changes the values of an individual to introduce new characteristics into the population [98].

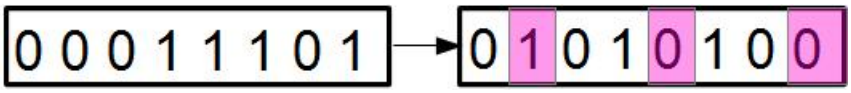


Figure 2.10: Mutation of a chromosome with the 1st, 4th, and 7th bits mutated.

Crossover can be generated from one of three different approaches: asexual, sexual, and multi-recombination [98]. The approach of the creation of the new generation based on crossover in these approaches is varied based on the number of parents used. In asexual, a single parent is used, while sexual uses two parents, as seen in Fig. 2.11. The multi-recombination approach uses more then two parents.

Using linear constraints to maintain the monotonic increasing functionality of the segments requires the crossover and mutation functions to be selected to maintain a feasible population to maintain those bounds.

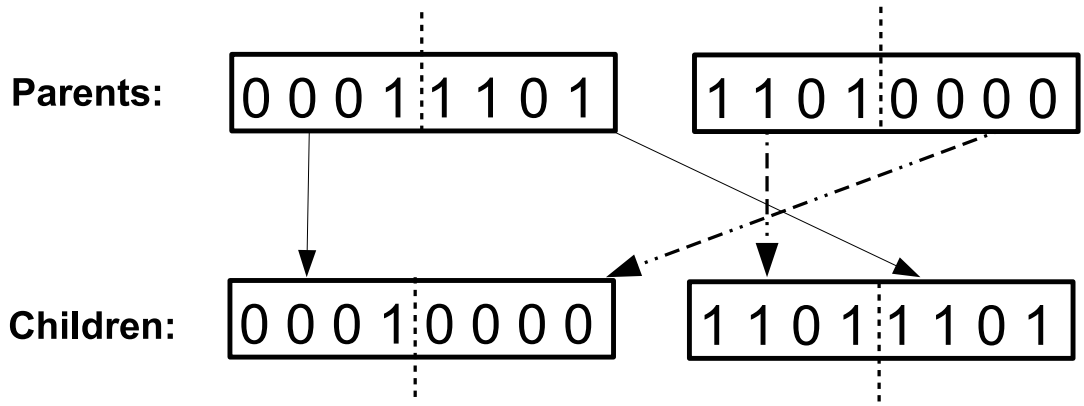


Figure 2.11: Crossover with two parents with the split point in the middle between the 3rd and 4th bit of the chromosome.

The overall process for a GA is as follows [102]:

Algorithm 1 Genetic Algorithm

```
1: procedure GA
2:     Initialize a population of N individuals
3:     Evaluate the fitness of the individuals
4:     while stopping conditions not met do
5:         Select individuals for reproduction probabilistically based on fitness
6:         Generate new individuals with crossover and mutation operators
7:         Evaluate the fitness of the new individuals
8:         Create next generation
```

The process continues until some stopping condition is met. Stopping criteria include an optima being locate or the maximum number of generations have occurred [102].

Genetic algorithms have been used in applications like optimizing the alphabet boundary placement for SAX [18, 54] and segmenting electrocardiogram (ECG) signals [16]. GA extends from the string/vector representation to become the basis of genetic programming when using tree representation [98]. Resource constrained scheduling problems have made used of GAs to optimize schedules [89]. The management of

groundwater resources has also benefited from the use of GAs with considerations into the pumping of water and the costs [103], so has the design of composite materials [104].

2.6 Oncology Domain Research

In the domain of oncology prognosis there has been a large amount of research using machine learning techniques [105] and statistical approaches [106]. Some common machine learning techniques used are BN [107], SVM [10, 86, 108, 109], artificial neural networks (ANN) [12, 109, 110], and decision trees (DT) [109, 111]. Commonly, statistical approaches are considered with oncology such as Cox Proportional Hazard [8, 112, 113, 114], Kaplan Meier [113, 115], and logistic regression [112, 114].

With the large amount of research that has been done in the oncology domain there are surveys [105, 116, 117] which establish machine learning and statistical approaches and features considered in the prognosis. In one survey [105], ANN, SVM, DT, semi-supervised learning (SSL), and BN approaches are created with accuracies greater then 70%. With the work presented it has a large number of limitations. These limitations include the number of categorical classifications, the patient populations size (only 4/18 used sets with more then 1000 patients, 6/18 had less then 100 patients considered with one only using a sample size of 31), and the limited data types, only

9/18 of the presented papers consider clinical data.

While the approaches in evaluation of oncology vary, so can the types of cancers and features available. Breast cancer often is the type in many studies [118]. Within the domain there are different types of data features such as clinical, genomic, images and demographic. Some studies look at features in different ways. One approach developed was to average the results of four different classifiers, one for each clinical, genomic, images and demographic data. These classifiers were implemented with techniques of ANN, BN, DT, NB, random forest (RF), and SVM [106]. Another study examined the SEER data set of 162,500 records using demographics and tumor data to comprise 16 features [109], this was evaluated for a 5 year binary classification with ANN (65%), SVM (51%) and SSL (71%).

Going beyond binary classification, one of the ways to accomplish multi-class classification is using an ensemble approach. Such an approach was done in a study of survival analysis with multiple classes of classification. There were multiple models developed to classify binary survival of 6 months (87% accurate), 1 year (80% accurate), 2 years (76% accurate), and greater then 2 years with the model being yes or no to that patient surviving beyond the classification time point [119]. Those models were based on 400 SVMs with linear kernels that then comprised an ensemble to determine survival. The survival result was an average of the SVM results. Instead of an ensemble approach a four class classification approach was done with the popular

Wisconsin Breast Cancer data set and using ANN [12].

With the time series data in oncology there is a probability for missing data or censored data. The managing of censored data has also been considered [120, 121]. Data can be censored for three reasons: (i) event of interest does not occur during the observation period of the study, (ii) the individual is lost prior to follow up, or (iii) the individual withdraws from the study [44].

Another common approach to considering prognosis is score or index based. One common method is the Nottingham Prognostic Index (NPI) that considers the tumor characteristics like the histological stage, lymph node stage, and tumor size [122]. Similarly in preparation for palliative care there is the Palliative prognostic index scale which considers other medical features, such as edema and delirium, that is used in the assessment of survival time with intervals of less then 3 weeks, between 3 and 6 weeks, and greater then 6 weeks [123]. In these cases a numerical value for a score is achieved based on the presence (or value) for a feature. In the scored which consider clinical values, the observations may be considered in real time or through summary approaches [124].

Chapter 3

Discrete Prognosis Using Bayesian Networks, Naïve Bayes and Support Vector Machines

Discrete, non-binary prognosis is explored using three classification techniques: Bayesian Networks, Naïve Bayes, and Support Vector Machines. The representation of the clinical data is considered through discrete interpretation and behavioral representation.

The Bayesian Network and Naïve Bayes results were presented in part at IEEE's Electro/Information Technology Conference 2016 and published in the proceedings [4].

3.1 Discrete Representation

The overall approach used for the discrete representation analysis is seen in Fig. 3.1. The baseline representation focused on discrete categorical representation and behavioral descriptors for two different data sets (see Section 3.1.1). For the classifier, four different approaches were considered to establish baselines for comparison: majority classifier, Bayesian Network (BN), Naïve Bayes (NB), and Support Vector Machines (SVM).

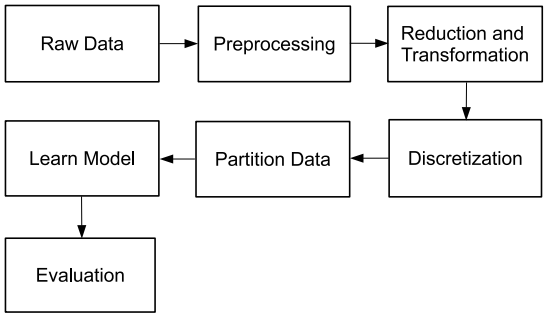


Figure 3.1: General process of each experiment.

For the classification approach, the core method which is used as the benchmark is the majority classifier. The utilization of BNs for classification was conducted as an expansive baseline with both two sets. The NB and SVM implementation was explored with one data set.

3.1.1 Data

Two data set, DS1 and DS2, comprised of nonuniform, multi-modal, time-limited time series samples were used. These data sets came from a private outpatient clinic and were provided by EMOL Health of Clawson, MI.

There were two types of observations available in the data sets: clinical and treatments. The observations for clinical data were for Albumin (ALB), Hemoglobin (HGB), and Weight (WT). Treatments observed included the administration dates for chemotherapy and two erythropoietins in both of the data sets; additionally DS1 contained observations of blood transfusion treatments.

The duration of data collection varies between patients depending on the number of visits. Observations are recorded over a maximum duration of two years. The determination of age at time of death was confirmed with the Social Security Death Index. For the two data sets used, the data set characteristics are described in Table 3.1. An example of the nonuniformity in observations of clinical data and administration of treatments is in Fig. 3.2.

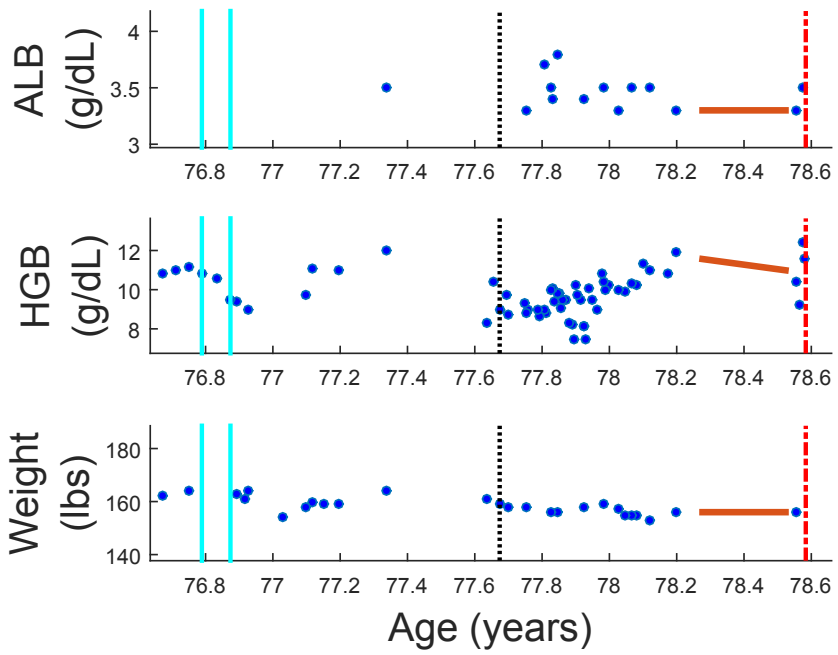


Figure 3.2: A sample from the data set, the treatments are shown along with a single difference trend. The vertical lines show administered treatments: solid (cyan) - erythropoietin, dashed (black) - blood transfusion, and dot-dashed (red) - chemotherapy. The brown line show the one difference trend (1 Diff). © 2016 IEEE [4].

3.1.2 Data Pre-Processing

The first step in preprocessing is the determination of a reference point for prediction of length of survival (LOS) for each sample. For any patient, the observations of each clinical attribute (ALB, HGB, WT) can have different final observation times, allowing for potential bias in selecting a LOS reference point or extrapolation errors in representing samples. In Fig 3.3 it shows the selection of the time t, the final observation over all types that ceases being measured first.

Table 3.1
DATA SET CHARACTERISTICS.

Properties	**DS1**	**DS2**
Patients, *num.*	1311	1922
Weight (WT) obs., *num.*	10,653	17,317
Albumin (ALB) obs. *num.*	5,547	9,545
Hemoglobin (HGB) obs., *num.*	17,481	27,895
Treatments, *num.*	3,411	3,440[a]
Age at death (yrs), *mean*	71.61	71.56
Age (yrs), *min/mean/max*	21.7 / 71.2 / 98.2	21.7 / 71.0 / 98.2
Obs./patient, *min/mean/max*	1 / 28.3 / 178	1 / 30.5 / 239
LOS from final obs. (days), *mean*	139	306

[a]. No blood transfusion data included.

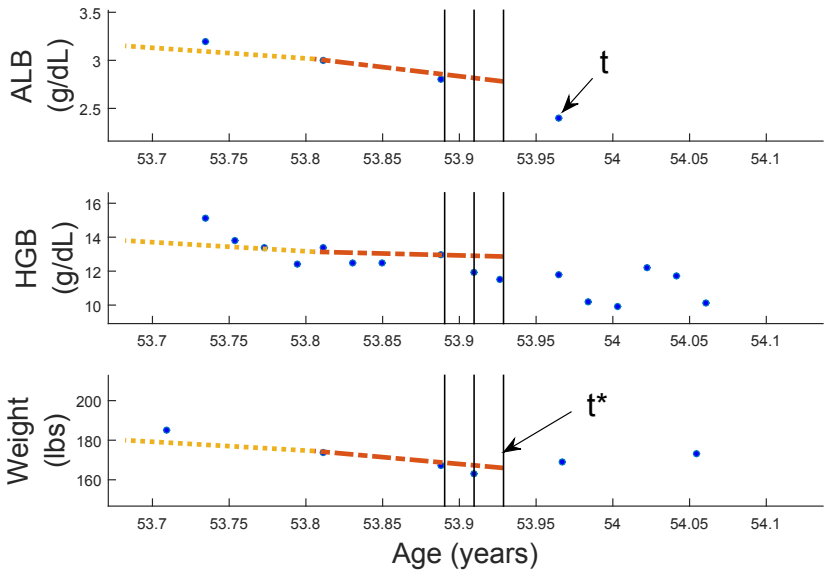


Figure 3.3: The solid (black) lines indicated where samples were estimated with a spacing of 7 days; two reference time points, t and t^*, are marked. The dotted (yellow) and dot-dashed (brown) lines show two difference trends (2 Diff). © 2016 IEEE [4].

To calculate the reference time point t^*, it is selected at random from a range about t for a patient. Two reference points are created to avoid bias, t_1^* and t_2^* are selected at random from a range about t, with $t_1^* \in [t-15, t+5]$ selected from the range of

15 days further from death to 5 days closer to death and $t_2^* \in [t-28, t+14]$. Only one reference point, t_1^* or t_2^* is used at a time for a given analysis.

From this data, discrete groupings for DS1 and DS2 follow medical based classification. Behavioral considerations of difference trends and splines will also be used (see Sections 3.1.3.3 and 3.1.3.4).

3.1.2.1 Data Set Reduction

In outpatient settings, the observations are made at nonuniform intervals. In the extreme, some clinical observations may be completely absent for a patient. The problem of missing data can be addressed by censoring. Typically, there are three causes for censoring: (i) an event of interest does not occur before the end of the study, (ii) an individual is lost to follow up, and (iii) an individual withdraws from the study [44].

In efforts to manage incomplete, absent or censored data for patients, filtering and cleaning techniques are used. Some approaches to cleaning the data sets can reduce the amount of patients by up to 50% [125]. To handle the occurrences of incomplete observations for patients in our study, two approaches are used. In the first approach, the input variables representing the patient's information are set to a specified category when there is no observation or realistic value for that observation available. In

the second approach, the patients who did not have at minimum one value for each of the clinical observation type were removed from the data set.

3.1.3 Data Representation

For the data sets, with the exception of chemotherapy, the clinical observations were made as a time series. The erythropoietin and blood transfusion observations are represented as a event occurrence count, taking the form of a single value which is then discretized. Several methods of representation of each patient's information are considered: numerical (treatments), sampling (see Section 3.1.3.1), and two different behavioral approaches (difference trends and splines, see Section 3.1.3.2).

All values will be discretized for evaluation. The discrete categories of the clinical observations are based on the general medical knowledge. Often clinical data has normal limits that may be dependent on gender; for this analysis, gender is unknown therefore, the maximum and minimum values across genders is used in the discretization. The numeric occurrences (number of treatments) are grouped into four categories. Patient age is discretized to achieve a uniform distribution across categories. Values of the boundaries are presented in Table 3.2.

For all input types, there is an invalid category. This discrete category is used when unrealistic values are found. It is also applied when the sample information does not

exist; a situation that may occur when there is no data at all or if the data requested is outside of the window of the patient's observed history.

Table 3.2
DISCRETE INPUT CATEGORIES. © 2016 IEEE [4].

Information Type	Units	Discrete Categories†
Albumin (ALB1, ALB2, ALB3)	g/dL	0-3.4,3.4-5.4, >5.4
Hemoglobin (HGB1, HGB2, HGB3)	g/dL	0-12, 12-17, 17-100
Weight (WT1, WT2, WT3)	lbs	0-116, 116-157, 157-188, 188-471
Chemotherapy Treatment (Chemo)		binary value significant of if chemotherapy was given to the patient
Blood Transfusion (Tcnt)	number of treatments	0, 1-4, 5-9, >9
Age at time t (age)	years	<59.4, 59.4-76.6, 76.6-82.2, >87.2,
Erythropoietin (Aranesp and Procrit) (Acnt and Pcnt)	number of treatments	0, 1-4, 5-9, >10,
Difference Trend (WTtr1, WTtr2, ALBtr1, ALBtr2, HGBtr1, HGBtr2)		Stable (< ± 0.1 g/dl, 2 lbs), Increasing, Decreasing
Splines (WTstr1, WTstr2, ALBstr1, ALBstr2, HGBstr1, HGBstr2)		<-0.5*st.dev., -0.5*st.dev. - 0.5*st.dev., >0.5*st.dev.,
Length of Survival (LOS)	days	0-21, 21-56, 56-168, >168

†All variables have classification categories for unrealistic and lack of observations.

3.1.3.1 Sampling Methods

Sampling occurs either from taking a value which exists or interpolating to find a value. Interpolation is used to extract all samples to create an uniform approach regardless of the presence of a sample or not at that specific time point. The first sample is taken at t^*, with successive samples taken at an interval of either 7 or 14 days. In Fig 3.3 three samples are shown with vertical lines at 7 day intervals.

3.1.3.2 Behavior Representations

The data of a patient is observed over time. As the data is time series, regardless of the nonuniform nature, there is an ability to describe behaviors of the data either over the entire series of specified sub-series. Two behavioral considerations are extracted from the data, difference trends and splines.

3.1.3.3 Difference Trends

Using the time point t^* as the first time point, the second time point is back 90 days. With these two time points the values at those times are estimated and then the

difference is determined.

$$trend = V_{t^*} - V_{t^*-90}. \tag{3.1}$$

From this value the classification of the behavior is defined as stable, increasing, or decreasing. The stable range specification is <0.1±g/dl for HGB and ALB, then <2 lbs for WT. This can be visualized on Fig 3.2. The threshold used to discretize the difference trend are estimated from the data set to result in an approximately uniform distribution of samples across the three behaviors.

A second variation considered is two behavioral trends, using the midpoint of the single trend (back 45 days from t^*). A two difference trend example is shown in Fig. 3.4. The trend is calculated for each segment. The trend descriptions follow a similar logic to work above for stable, increasing, and decreasing in each of the two sub-series [126].

3.1.3.4 Splines

Splines are used to describe the behavior of the observations. A two-piece second order spline is used to fit the entire observation period for ALB, HGB, and WT observations for a patient (unlike the difference trend which has a recent specified period of considerations); see Fig. 3.5. The splines' slope coefficient is used as the input to predict LOS.

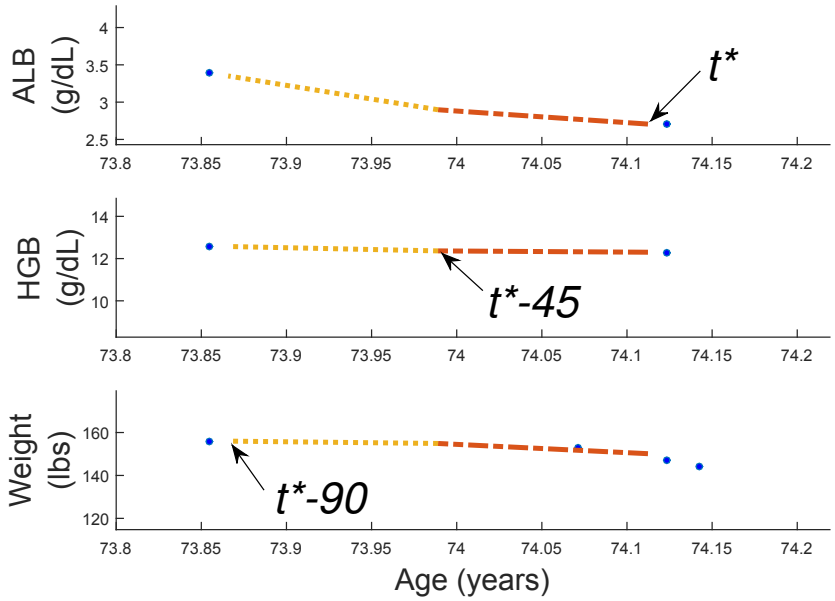


Figure 3.4: Two difference trends with the knot points shown. The first trend is the brown dot-dash line from t^* back 45 days and the second trend segment is the yellow dotted line for $t^* - 45$ days to $t^* - 90$ days.

3.1.3.5 Summary of Data Representations

There are multiple ways discussed to represent the patient observations: clinical data samples, difference trends, splines. Different combinations of representations will be considered in the evaluation, and the options available are summarized in Table 3.2. For example, the number of clinical data samples considered varies from zero to three. The number of difference trends included in the evaluation is zero to two. The spline information is either included or not.

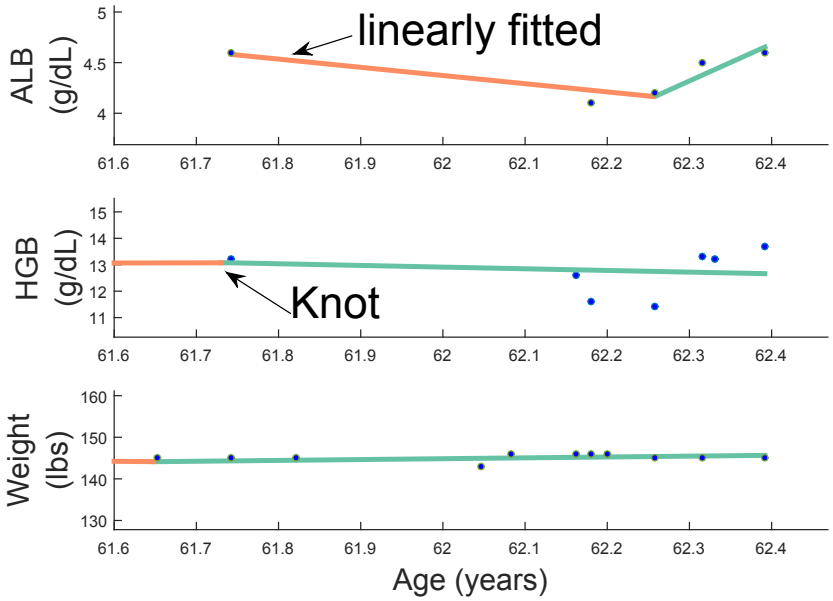


Figure 3.5: Example of the fitting of splines for a sample with the nonuniform, multi-modal, time-limited time series property, the orange and green lines illustrate the two-piece splines.

3.2 Experimental Design

For any particular data set, DS1 or DS2, and its representation approach, then four classification methods were considered for predicting LOS: the majority classifier, Bayesian networks, Naïve Bayes, and SVMs. The data was split to training and test sets with a ten-fold cross validation approach. A t-test was utilized to confirm there is a statistical difference for LOS prediction, producing a p-value. When the p-value is less then 0.05 the difference can be seen as statistically significant.

3.2.1 Majority Classifier

A majority classifier was used as the baseline in this evaluation. The first component of this approach is the determination of the number of patients in the training set that have the specified length of survival (LOS) class, y_i where $y_i \in \mathbf{Y}$, this is done for all classes. The majority class i is selected to be the class y_i with the largest proportion of the training data. This class label y_i is then used as the predicted LOS for the test set.

3.2.2 Bayesian Network Development

The network structure for BN initially includes all data representations available in the analysis, shown in Fig. 3.6. Depending on what combination of representations (number of occurrences, samples, differences, and splines) that are being considered the network structure is modified to fit those variables. The change in the representations considered is used for testing the impact on the LOS prediction. For all networks not involving the full combination of representations, any variables excluded require removing the associated nodes and all incident edges in the network, then reconnecting the broken paths. An example of a network that does not involve the direct clinical observations through sampling or spline behavior is Fig. 3.7, that also

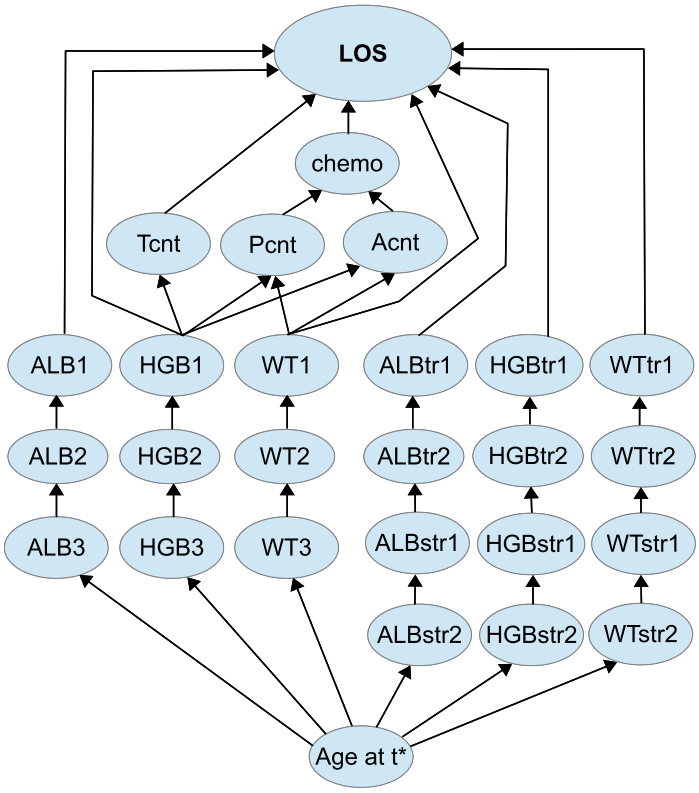


Figure 3.6: Full Network. © 2016 IEEE [4].

happens to be the configuration with the best accuracy observed during evaluations.

Within each fold of the cross-validation, the parameters of the network are learned through Expectation Maximization (EM) within the built-in functions of the GeNIe software package [75, 76]. The test data is evaluated on the trained network in GeNIe, where the LOS category with the maximum probability is selected, then analyzed with the mean and standard deviation across the folds.

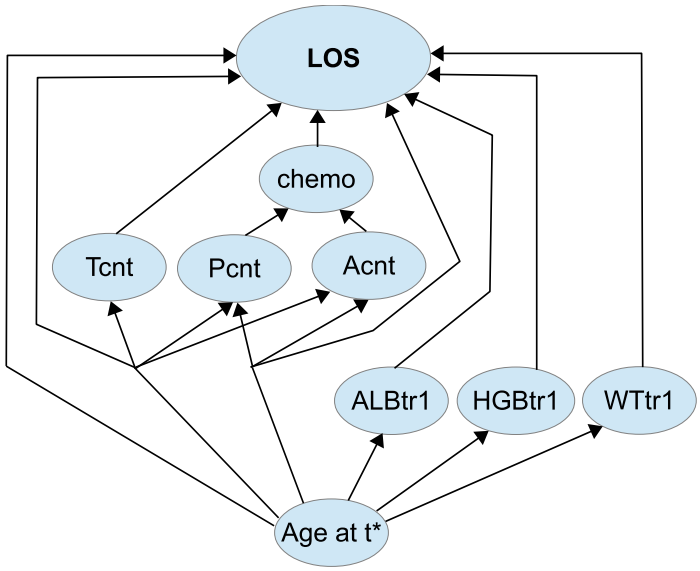


Figure 3.7: Network configuration with best accuracy.

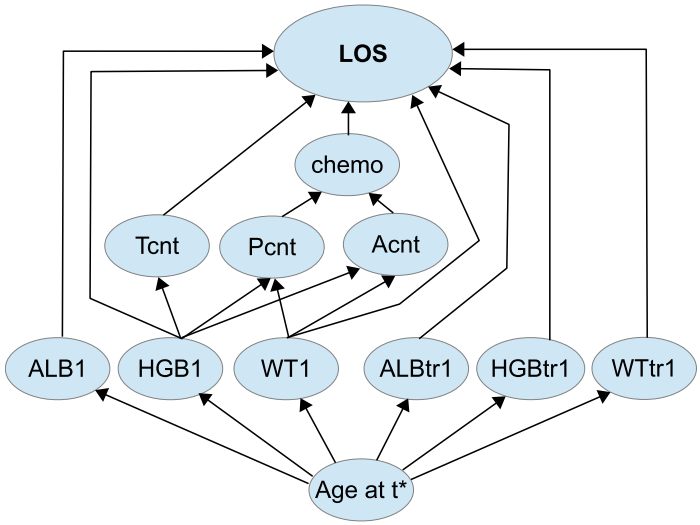


Figure 3.8: Network configuration with worst accuracy.

3.2.3 NB and SVM Classifiers

Naïve Bayes, as discussed in 2.3.1, and SVMs, discussed in 2.3.3, were modeled and evaluated in Matlab. There are methods which exist for the parameter selection such

as nested cross validation, particle swarm optimization [127] or sequential minimization optimization [128]. The approach chosen is to utilize a nested cross validation to select the parameters. For the parameter selection of the SVM RBF Kernel, five different options were selected for σ and four for C based on literature [129, 130, 131, 132]. The selected parameters for RBF to consider are $\sigma = \{0.001, 0.01, 0.1, 1, 2\}$ and $C = \{0.1, 1, 10, 100\}$. For the linear kernel, the same cost parameters are considered.

3.3 Results

The results are presented for the two data sets for the BN, DS1 and DS2, with conclusions and comparisons to follow. NB and SVM were evaluated with only DS1. Since a doctor's prediction of LOS is not available for this study, a majority classifier was used as the baseline; that is, the LOS category that had the most patients in the training set was selected as the predicted LOS for all patients in the test set.

3.3.1 DS1 Results from Evaluation with Bayesian Networks

Initially, the results on DS1 focused on a representation with one difference trend and the spline information included; the clinical data samples spacing considered was

7 or 14 days with the number of clinical data samples going from 0 to 3 and both time reference points examined, t_1^* and t_2^*. The results are seen in Table 3.3, which shows that the BN prediction of LOS outpaces the majority classifier in all cases (p-values <0.05). When comparing the performance between the predictions using t_1^* and t_2^* there is no statistical difference (p-values of 0.36-0.64), but t_2^* results in higher accuracies in all cases. Therefore, t_2^* will be the LOS reference point method presented in further examinations. Also, when samples are included the accuracy is reduced, but beyond the first sample, the number of samples included does not show any effect on the accuracy.

When examining the effect of the number of clinical data samples to include, there is no difference in the accuracy between 1, 2, or 3 samples (all of the pairwise comparison have a p-value of 1.000). The difference between having 0 or 1 clinical data samples included in the evaluation is $\approx 5\%$ greater for no samples. This difference is statistically significant (p-value of 0.002).

In Table 3.4, the LOS prediction results with the inclusion (or not) of spline and Diffs variables are compared for t_1^* and t_2^* respectively. In general, the conclusion drawn from these results are similar (with p-values >0.2), and that the inclusion of clinical data samples has a negative impact on the prediction results. The inclusion of any trend or spline information improves the accuracy of LOS prediction over the majority classifier. When examining the inclusion of spline information, for both t_1^*

Table 3.3
DS1 LOS Prediction Accuracy (Acc.) using Bayesian Networks with 1 Diffs and spline information included.

Num of Samples	Spacing (days)	Acc. (%) - p-value t_1^*	p-value‡	t_2^*	p-value‡
0	7	**35.7±3.2**	<0.001	**37.1±2.3**	<0.001
1	7	31.4±3.9	0.012	32.1±4.5	0.045
2	7	31.4±3.9	0.012	32.1±4.5	0.045
3	7	31.4±3.9	0.012	32.1±4.5	0.045
0	14	**35.7±3.3**	<0.001	**37.1±2.3**	<0.001
1	14	31.4±3.9	0.012	32.1±4.5	0.045
2	14	31.4±3.9	0.012	32.1±4.5	0.045
3	14	31.4±3.9	0.012	32.1±4.5	0.045

‡Comparison with the majority classifier produced the p-value.

Table 3.4
DS1 LOS Prediction Accuracy (Acc.) using Bayesian Networks†.

Num of Samples	Num of Diffs	Acc. (%) Spline	p-value‡	Acc. (%) No Spline	p-value‡
0	0	33.9±4.4	<0.001	Maj. Class	—-
0	1	**35.7±3.0**	<0.001	35.7±3.3	<0.001
0	2	35.5±5.9	0.0019	35.5±5.9	0.0019
1	0	34.2±4.2	<0.001	34.8±3.3	<0.001
1	1	29.7±2.1	0.003	31.4±3.9	0.005
1	2	32.3±4.6	0.007	32.3±4.6	0.007
0	0	36.3±3.7	<0.001	Maj. Class	—-
0	1	**37.1±2.3**	<0.001	**37.1±2.3**	<0.001
0	2	33.5±1.7	<0.001	33.5±1.7	<0.001
1	0	34.6±1.6	<0.001	33.9±3.3	<0.001
1	1	32.1±4.5	0.045	32.1±4.5	0.045
1	2	32.6±3.0	0.002	32.7±3.0	0.002

† Reference point t_1^* results are presented above the triple line and t_2^* below.
‡Comparison with the majority classifier produced the p-value.

and t_2^*, with the case when there are no clinical observations included, there is no difference observed. The remaining combinations have a slight difference in accuracy, but they are not statistically significant in most cases, the few cases of difference are

Table 3.5
DS1 LOS Prediction Accuracy (Acc.) using Bayesian Networks
Cases of statistical differences.

DS	t^*	Constants	Condition Varied	p-value
1	1	1 Sample no Spline	0 vs 1 Diffs	0.005
1	2	0 Sample, Spline and no Spline	1 vs 2 Diffs	<0.001
2	1	1 Sample, no Spline	0 vs 1 Sample	0.047
2	1	0 Sample, no Spline	0 vs 2 Diffs	0.013
2	1	1 Sample, Spline	0 vs 1 Diffs	0.024
2	1	1 Sample, Spline	0 vs 2 Diffs	0.049
2	1	1 Sample, no Spline	0 vs 2 Diffs	<0.001

presenting in Table 3.5. Comparing the possible combinations of factors in the BN there were only three cases which had differences that could be considered statistically different. When there is one sample on t_1^* between the no trend and 1 trend (p-value 0.005), the case between 1 and 2 trends with splines (p-value 0.07), then between the 1 or 2 trends independent of spline presence with no samples for t_2^* (p-value <0.001).

While the influence of difference trends over the majority has been shown, the impact of the different number of trends is less significant. Comparing the accuracies between the 0, 1, and 2 difference trends, the accuracy is within 4%. For the case of no samples, the highest accuracy was observed with one difference trend. The result is statistically significant over the model with 2 difference trends (p-value <0.001). The results comparing 0 and 2 difference trends is also significant (p-value 0.013). The higher accuracy of one difference trend compared to none, is not significant (p-value 0.517).

Table 3.6
DS1 LOS Performance Analysis using Bayesian Networks†.

Model	**Overall Accuracy (%)**	**Longer Prediction(%)**	**Shorter Prediction(%)**
Spline, 1 Diffs	35.70	27.38	36.92
Spline Only	33.87	28.14	37.99
1 Diffs Only	35.70	27.38	36.92
Spline,1 Diffs	37.07	27.15	35.77
Spline Only	36.31	30.21	33.48
1 Diffs Only	37.07	27.15	35.77

†Above the triple line is based on reference point t_1^* and below on t_2^*.

Overall, the highest accuracy is when there is no clinical data samples, a 1 Diffs and no spline for both t_1^* and t_2^*, with accuracies of 35.7±3.3 and 37.1±2.3 respectively and both with p-value <0.001 compared to majority classifier.

In Table 3.6, the percentage of patients whose LOS is classified correctly, predicting a longer LOS than the actual, and predicting a shorter LOS are shown for three models. These models are the cases where there is only the splines and 1 Diffs, the splines only and then the 1 Diffs only; no clinical data samples were considered in these models. For the incorrect classifications, they tend to be pessimistic, predicting a shorter LOS. Also, the predictions are typically only off by one category, e.g. a sample with a LOS of 21-56 days, would be incorrectly classified as less than 21 days. In an objective view, if an error is to be made the family of a terminal patient would typically prefer the longer survival.

3.3.2 Reduced DS1 Evaluation with Bayesian Networks

There is a known bias in the evaluation of the data sets due to missing observations occurring for some patients. Filtering of the data removes patients who do not have at least one observation of each type of clinical measurements (ALB, HGB, and WT). For example, a patient who has never had their weight recorded would be removed from consideration. When these patients were removed in DS1, the 10-fold cross validation test sets dropped from 131 to 63 patients, a the population saw a reduction of 52% (48% of the population is retained, 639 patients).

The reduce set has folds created as done for the full DS1, however the distribution of patients into folds are different to maintain uniform behaviors between folds (instead of to the larger set). For the reduced set, the majority classification has an accuracy of 30.8±0.7. Overall, the results are often on par with the majority classifier. In terms of the behavior, or identification of which data representations lead to accurate LOS predictions, the results are similar to the non-reduced data set. For example, clinical data samples does not improving the accuracy, inclusion of spline information does not affect the performance, and the predicted outcome maintains the tendency for shorter LOS when not correct. A change from the full data set is that having one difference trend does not improve performance compared to no difference trends included. Using the reduced set, the best model had an accuracy of 32.7±3.7 which is

not significantly better than the majority classifier (p-value 0.174). The performance of the BN models on the reduced data sets is overall lower compared to the full data set. This is not an unexpected result, in part due to smaller patient population sizes greatly affecting the BN learning algorithms, specifically the conditional probability table (CPT) estimates.

Bias is an expected factor when dealing with data sets which contain incomplete and missing records for patients. In order to see the impact of these situations, patients who do not contain at least one observation of each type of clinical measurements (ALB, HGB, and WT) are filtered. For example, a patient who has never had their WT recorded would be removed from consideration. When these patients were removed in DS1, the 10-fold cross validation test sets dropped from 131 to 63 patients, a the population saw a reduction of 52% (48% of the population is retained, 639 patients). Reduction of population does effect the EM learning, which leads towards lower accuracies.

Overall, the results are often on par with the majority classifier for the filtered set. In terms of the behavior, or identification of which data representations lead to accurate LOS predictions, the results are similar to the non-reduced data set. For example, clinical data samples does not improving the accuracy, inclusion of spline information does not affect the performance, and the predicted outcome maintains the tendency for shorter LOS when not correct. A change from the full data set is that having 1

Diffs does not improve performance compared to no Diffs included. The performance of the BN models on the reduced data sets is overall lower compared to the full data set. This is not an unexpected result, in part due to smaller patient population sizes greatly affecting the BN learning algorithms, specifically the CPT estimates.

3.3.3 DS2 Evaluation with Bayesian Networks

The second data set, DS2, is analyzed to confirm or validate the behaviors seen in DS1. Once again, the majority classifier is used as a baseline; for this data set the majority classifier has an accuracy of 43.7±0.2 for t_1^* and t_2^*. Note, this value is much larger than that of DS1.

The trend behaviors were examined just as was done with DS1. The accuracies in Table 3.7, show the best performance occurring when the network includes only the spline data (47.0±1.5 - t_1^* and 47.9±1.8 - t_2^*). The inclusion and increasing number of Diffs decreases the prediction accuracy for both t_1^* and t_2^*.

The lowest performing network is the same as in DS1, when the network contains 2 Diffs; this behavior is independent of the inclusion of splines and clinical data samples.

For t_1^* and t_2^*, beyond the comparison with the majority classifier there are few cases when compared that have statistical difference. The cases are presenting in Table 3.5

Table 3.7
DS2 LOS Prediction Accuracy (Acc.) using Bayesian Networks†.

Num of Samples	Num of Diffs	Acc. (%) Spline	p-value‡	Acc. (%) No Spline	p-value‡
0	0	47.0±1.5	<0.001	Maj. Class	—-
0	1	46.0±2.1	0.006	46.0±2.1	0.006
0	2	43.9±2.3	0.739	43.9±2.3	0.739
1	0	43.6±4.2	0.984	**47.5± 3.5**	0.007
1	1	41.8±3.7	0.123	41.8±3.7	0.123
1	2	39.8±3.3	0.004	39.8±3.3	0.004
0	0	**47.9±1.8**	<0.001	Maj. Class	—-
0	1	45.6±1.7	0.007	45.6±1.7	0.007
0	2	44.3±1.6	0.266	44.3±1.6	0.266
1	0	45.6±2.8	0.062	44.9±1.0	0.007
1	1	43.1±2.9	0.562	43.1±2.9	0.562
1	2	39.7±3.1	0.003	39.7±3.1	0.003

†The LOS reference point t_1^* is above the triple line and t_2^* is below.
‡Comparison with the majority classifier produced the p-value.

for t_1^*, t_2^* has statistical differences observed for all of the combinations with p-values <0.025, except for when there is 1 sample between 0 and 1 Diffs.

For t_1^* beyond comparison with the majority classifier, the only occurrences of statistical difference when comparing models results in comparison are with one sample no difference trend (p-value 0.047), zero samples and 0 vs 2 difference trends (p-value 0.013), for 1 sample there are statistical differences for 0 vs 1 difference and a spline (p-value 0.024), 0 vs 2 trend with spline (p-value 0.049), and most significantly different between 0 and 2 trends with no spline (<0.001).

The statistical differences between combinations on t_2^* are p-values <0.018 for combinations of 0 samples comparing different trend inclusion, for both cases of splines

consideration. With one sample considered all combinations excluding 0 versus 1 difference trends with a spline has p-values <0.025.

Prognostic tendencies were looked at on the same three conditions that were done on DS1. These models are the cases where the BN contains with the splines and one difference trend included, only the splines, and then 1 Diffs only; no clinical data samples were considered in these models. The higher overall accuracies reduces the potential for longer and shorter LOS predictions. In Table 3.8, the models are presented for their prediction distribution. The likely outcome when the prognosis is incorrect is a shorter LOS (consistent with DS1); this is supported by the shorter category being the incorrect outcome more then twice as frequent as the longer prediction. This is not unexpected due to LOS distribution having so many patients with LOS of longer then 168 days. The short prediction is consistent with the DS1 behavior, which suggest the bias from the data set distribution is not the only influence on this tendency.

Table 3.8
DS2 LOS Performance Analysis using Bayesian Networks†.

Model	**Overall Accuracy (%)**	**Longer Predicted(%)**	**Shorter Predicted(%)**
Spline,1 Diffs	46.04	16.03	37.93
Spline Only	46.99	11.36	41.65
1 Diffs Only	46.04	16.03	37.93
Spline, 1 Diffs	45.63	17.23	37.14
Spline Only	47.93	9.90	42.17
1 Diffs Only	45.63	17.23	37.14

†Above the triple line is based on reference point t_1^* and below on t_2^*.

3.3.4 Evaluation with Naïve Bayes

Utilization of the established representation used with the BNs was considered with NB. The general trends of performance remained the same as with the BN and are shown in Table 3.9 for the NB evaluation.

Table 3.9
DS1 LOS Prediction Accuracy (Acc.) using Naïve Bayes†.

Num of Samples	**Num of Diffs**	**Acc. (%) Spline**	**p-value‡**	**Acc. (%) No Spline**	**p-value‡**
0	0	28.7±3.7	0.610	Maj. Class	—-
0	1	33.0 ±2.6	<0.001	35.0 ±4.1	0.001
0	2	32.4 ±4.6	0.014	36.2 ±4.5	<0.001
1	0	32.9±3.2	<0.001	34.0±3.6	<0.001
1	1	33.4±2.9	<0.001	**36.8±3.8**	<0.001
1	2	32.9±3.2	<0.001	34.0 ±3.6	<0.001
0	0	32.0±3.8	0.011	Maj. Class	—-
0	1	33.5±5.2	0.015	35.9±5.2	0.002
0	2	32.1±4.6	0.039	32.9±5.2	0.032
1	0	33.0±4.1	0.005	35.3±2.9	<0.001
1	1	**37.9±3.8**	<0.001	36.4±4.3	<0.001
1	2	31.9±4.0	0.028	34.1±3.1	<0.001

†Reference point t_1^* results are presented above the triple line and t_2^* below.
‡Comparison with the majority classifier produced the p-value.

3.3.5 Evaluation with Support Vector Machines

Rounding out the discrete LOS classification approaches, two different kernels were used with support vector machines (SVMs): linear and RBF. The overall trend of benefits from including behavioral representation continued as seen with the linear kernel in Table 3.10 and the RBF in Table 3.11.

Table 3.10
DS1 LOS Prediction Accuracy (Acc.) using SVM- Linear†.

Num of Samples	**Num of Diffs**	**Acc. (%) Spline**	**p-value‡**	**Acc. (%) No Spline**	**p-value‡**
0	0	37.4±2.5	<0.001	Maj. Class	—-
0	1	37.6±2.7	<0.001	34.4±2.8	<0.001
0	2	**39.1±3.3**	<0.001	36.00±2.6	<0.001
1	0	37.7±2.7	<0.001	34.25±3.2	<0.001
1	1	39.1±4.3	<0.001	35.47±3.1	<0.001
1	2	38.7±3.0	<0.001	35.77±2.6	<0.001
0	0	35.8±3.0	<0.001	Maj. Class	—-
0	1	36.4±4.2	<0.001	33.9±3.1	<0.001
0	2	38.1±3.1	<0.001	35.5±3.3	<0.001
1	0	38.1±3.1	<0.001	34.9±2.1	<0.001
1	1	38.1±3.6	<0.001	35.8±3.1	<0.001
1	2	**38.4±3.8**	<0.001	36.2±2.0	<0.001

†Reference point t_1^* results are presented above the triple line and t_2^* below.
‡Comparison with the majority classifier produced the p-value.

The impact on the number of samples considered is seen in Table 3.13 for the linear kernel and Table 3.12 for the RBF.

Table 3.11
DS1 LOS Prediction Accuracy (Acc.) using SVM- RBF†.

Num of Samples	Num of Diffs	Acc. (%) Spline	p-value‡	Acc. (%) No Spline	p-value‡
0	0	36.2±3.7	<0.001	Maj. Class	—-
0	1	33.6±2.5	<0.001	34.1±3.7	<0.001
0	2	34.3±3.1	<0.001	33.9±4.6	<0.001
1	0	34.5±4.1	<0.001	34.6±4.2	<0.001
1	1	32.7±3.3	<0.001	33.3±4.2	<0.001
1	2	33.6±3.5	<0.001	34.1±3.5	<0.001
0	0	35.2±3.2	<0.001	Maj. Class	—-
0	1	34.3±3.9	<0.001	35.3±3.1	<0.001
0	2	34.3±2.5	<0.001	35.0±3.3	<0.001
1	0	34.3±1.9	<0.001	33.3±2.2	<0.001
1	1	33.0±1.8	<0.001	34.7±3.0	<0.001
1	2	33.0±1.8	<0.001	34.7±3.0	<0.001

†Reference point t_1^* results are presented above the triple line and t_2^* below.
‡Comparison with the majority classifier produced the p-value.

Table 3.12
DS1 LOS Prediction Accuracy (Acc.) using SVM with RBF Kernel†.

# of Diffs	Splines	# Samples	%Acc.	
			7day	**14**
1	no	0	34.1±3.7	34.1±3.7
1	no	1	33.3±4.2	34.0±4.1
1	no	2	33.3±4.2	33.3±4.4
1	no	3	**34.2±3.8**	33.0±4.1
1	yes	0	33.6±2.5	33.6±2.5
1	yes	1	32.7±3.3	32.7±3.3
1	yes	2	33.0±3.3	32.9±3.3
1	yes	3	32.6±2.8	32.7±3.3
1	no	0	35.3±3.1	35.3±3.1
1	no	1	34.7±3.0	**35.6±3.8**
1	no	2	35.5±3.1	35.4±3.3
1	no	3	35.0±2.8	35.1±3.2
1	yes	0	34.3±3.9	34.3±3.8
1	yes	1	33.0±1.7	33.3±3.8
1	yes	2	33.6±4.0	33.6±4.5
1	yes	3	33.3±4.2	33.4±4.5

†Reference point t_1^* results are presented above the triple line and t_2^* below.

Table 3.13
DS1 LOS Prediction Accuracy (Acc.) using SVM with Linear Kernel†.

# of Diffs	Splines	# Samples	Acc. (%)	
			7 Days	**14 Days**
1	no	0	34.4±2.8	34.4±2.8
1	no	1	35.5±3.1	35.5±3.0
1	no	2	37.2±2.8	37.8±2.3
1	no	3	37.8±2.6	37.8±2.9
1	yes	0	37.6±2.7	37.6±2.7
1	yes	1	39.1±4.3	39.1±4.3
1	yes	2	38.5±4.2	39.6±3.7
1	yes	3	37.3±4.1	**39.7±3.5**
1	no	0	33.9±3.1	33.9±3.1
1	no	1	35.8±3.1	35.8±3.1
1	no	2	37.4±4.1	36.9±3.8
1	no	3	37.3±4.1	36.8±3.8
1	yes	0	36.4±4.2	36.4±4.2
1	yes	1	38.1±3.6	38.1±3.6
1	yes	2	37.3±3.2	37.8±4.0
1	yes	3	**38.1±3.4**	37.5±3.9

†Reference point t_1^* results are presented above the triple line and t_2^* below.

3.4 Summary

Through the evaluations the establishment of baselines to be considered as benchmarks to compare with in future work has been conducted. In Table 3.14 the summary results are presented.

The improvement of accuracy observed with the BN, NB, and SVM compared with the majority classifier allows for the consideration of more behavioral aspects, this considers behavioral representation as solution to nonuniform data. With this conclusion, other approaches to behavioral representation should be considered this can take a form closer to the original data as planned with the Piecewise Aggregate Approximation (PAA) work or look at modeling events with methods like temporal abstraction [63, 126, 133].

The performance trends were different from the BN with the NB due to the best performing configurations having clinical samples, this suggests that more considerations be made to incorporating the clinical observations, this supports the decision to explore PAA.

There is no comparative analysis made to work in the literature due to the uniqueness of the data set and the focus being on the representation with the impact determined by the length of survival classification analysis.

Table 3.14
Summary of important results.

Method	Representation Parameters	Acc. (%)
Maj. Classifier	t_1^*, DS1	28.6±0.4
Maj. Classifier	t_2^*, DS1	27.4±0.2
Maj. Classifier	t_1^*, DS2	43.7±0.2
Maj. Classifier	t_2^*, DS2	43.7±0.2
BN	t_1^*, DS1, No Samples, 1 Diffs†	35.7±3.3
BN	t_2^*, DS1, No Samples, 1 Diffs†	37.1±2.3
BN	t_1^*, DS2, 1 Sample, No Diffs, No Splines	47.5±3.5
BN	t_2^*, DS2, No Samples, No Diffs, Splines	47.9±1.8
NB	t_1^*, DS1, 7 day, 3 Samples, Diffs, Splines	37.5±3.9
NB	t_1^*, DS1, 14 day, 3 Samples, 1 Diffs, Splines	38.2±3.3
NB	t_2^*, DS1, 7 day, 3 Samples, 1 Diffs, Splines	37.9±3.8
NB	t_2^*, DS1, 14 day, 3 samples, 1 Diffs, Splines	38.2±3.3
SVM-Linear	t_1^*, DS1, 7day, No Samples, Diffs, Splines	39.1±4.5
SVM-Linear	t_1^*, DS1, 14 day, 3 Samples, 1 Diffs, Splines	39.7±4.3
SVM-Linear	t_2^*, DS1, 7 day, 3 Samples, Diffs, Splines	38.7±4.2
SVM-Linear	t_2^*, DS1, 14 day, 2 Samples, 1 Diffs, Splines	38.8±4.2
SVM-RBF	t_1^*, DS1, 7 day, 0 Samples, No Diffs, Splines	36.2±4.6
SVM-RBF	t_1^*, DS1, 14 day, 2 Samples, No Diffs, No Splines	36.5±4.6
SVM-RBF	t_2^*, DS1, 7 day, 2 Samples, 1 Diffs, No Splines	35.5±4.2
SVM-RBF	t_2^*, DS1, 14 day, 0 Samples, No Diffs, Splines	35.8±4.5

† For both Spline and No Spline included.

Chapter 4

Regression Analysis with Behavioral Representation

The work in this chapter is published in part in Advances in Science, Technology, and Engineering Systems Journal special edition issue on Recent Advances in Electrical and Electronics Engineering [32].

The limitations of discrete length of survival (LOS) prognostic classification are in the finite number of classes representing the different LOS categories. Rather then be restricted to the classes, the next logical phase is regression analysis to be able to create a more detailed prognosis of survival. The format of the representation of

the input variables to the models was maintained from the discrete category exploration with the exception that values were not discretized prior to analysis they were standardized.

4.1 Methodology of Evaluation Techniques

Regression has numerous techniques available for consideration. The methods selected for this analysis were Linear Regression (Linear), Quadratic Regression (Quadratic), Gaussian Process (GP) with a constant basis, and Support Vector Regression (SVR) using the radial basis function (SVR-RBF) and linear (SVR-lin) kernels.

4.2 Experimental Design

The data was sampled and interpreted as in Section 3.1, but in place of the discretization for the clinical samples a zero mean, unit variance standardization was used. The combination of zero to six samples were considered in the evaluation. The behavioral representations considered are splines and both one difference trend (1 Diffs) and two difference trends (2 Diffs). The treatments (blood transfusion, chemotherapy and erythropoietins) are also considered as inputs.

Cross validation was used with all methods and nested cross validation for SVR parameter selection. The cost parameters available for both the linear and the radial basis function (RBF) kernels were $C = \{0.1, 1, 10, 50, 100, 500\}$. The sigma values of the RBF considered the values $\sigma = \{0.1, 1, 2, 5, 10\}$. For statistical verification, the t-test was used to produce p-values.

The evaluation of the regression models used two approaches, the absolute and the relative error. Absolute error is measured in the days, and relative error is a percentage with 1 being 100%.

The data set will be evaluated on in full and reduced form to see the impact of missing observations on the performance. The reduction of the data set population will be done to remove patients with clinical data types that lack observations entirely. This approach is explained in further detail when it was established in previous work (see Section 3.1.2.1).

4.3 Results

In the evaluation many different models were evaluated with the 5 different regression approaches. Starting out the comparison, complete patient population data set was utilized and compared against the reduced population data set (see Table 4.1 for the

complete version and Table 4.2). Additional results from the models not presented here can be found in Appendix C.

Table 4.1
ONE SAMPLE WITH t_2^* AND 14 DAY SAMPLE SPACING AND COMPLETE DATA.

# of Diffs	Splines	Median Relative Error				
		SVR-lin	**SVR-RBF**	**Linear**	**Quadratic**	**GP**
0	0	0.737	0.791	1.030	1.025	1.058
0	1	0.729	0.810	0.889	1.013	0.993
1	0	0.737	0.8135	1.022	1.053	1.081
1	1	0.734	0.837	0.896	1.146	**0.977**
2	0	**0.715**	**0.813**	**0.980**	**0.982**	1.064
2	1	0.727	0.824	0.888	1.115	1.015

In comparing the benefit of the Diffs and splines, one sample was considered in Table 4.2 for each of the regression methods. The complete data set showed relative errors with the SVR approaches between 72 and 84%, with the quadratic regression exceeding 100% error. The best preforming regression models noted in Table 4.1 occurred with the one sample, two Diffs and no splines for all regression approaches except for the Gaussian Process that performed best with one Diffs and splines.

With the outpatient setting, there are patients which lacked observations of some clinical sample types entirely. This approach used is the same as outlined in Section 3.1.2.1. By removing those patients the data set became less sparse since there was less regions to be filled in with zeros, the results in turn improved.

The same analysis can be done for the reduced data set. The best performance was

observed with one trend and no splines with the SVR having a linear kernel. The worst performance was consistently observed with the quadratic regression. The SVR with the RBF kernel performed the best with no Diffs or splines, followed by the condition of one trend and independent of splines. The linear regression best performed with the splines and one or no Diffs. The Gaussian Process best performed with the splines and no Diffs. In summary, this shows the presence of some temporal representation in conjunction with a sample to have the best results. It also favors the SVR methods and linear regression in the evaluation.

Table 4.2
ONE SAMPLE WITH t_2^* AND 14 DAY SAMPLE SPACING AND REDUCED DATA.

# of Diffs	Splines	Median Relative Error				
		SVR-lin	SVR-RBF	Linear	Quadratic	GP
0	0	0.636	**0.819**	0.881	**0.958**	0.874
0	1	0.649	0.844	**0.800**	1.064	**0.870**
1	0	**0.619**	0.834	0.8845	0.9796	0.874
1	1	0.634	0.834	0.803	1.182	0.879
2	0	0.720	0.868	0.866	0.979	0.878
2	1	0.665	0.877	0.818	1.330	0.893

Looking at both the complete (Table 4.1) and reduced data (Table 4.2 set performances. The relative error differences between zero samples and more samples shows a greater fluctuation in the performance. When there are no Diffs or splines there is up to 1.55% difference in relative error, with the best case having 2 samples. The performance overall is improved with some samples and Diffs and splines. The best being with no Diffs, with the splines and zero or one samples. These results are seen in Appendix C. The only difference observed was with the reduced set the error was

Table 4.3
MEDIAN RELATIVE ERROR FOR t_2^* FOR DATA REPRESENTATIONS OF 1 DIFFS, SPLINES, AND VARIOUS NUMBER OF CLINICAL SAMPLES FROM REDUCED DATA SET WITH 14 DAY SAMPLE SPACING.

	Median Relative Error				
Samples	**SVR-lin**	**SVR-RBF**	**Linear**	**Quadratic**	**GP**
0	0.658	**0.778**	0.838	**1.011**	**0.860**
1	0.634	0.834	**0.800**	1.182	0.879
2	0.631	0.828	0.830	1.257	0.928
3	0.631	0.880	0.811	1.425	0.933
4	0.655	0.817	0.834	1.686	0.900
5	**0.630**	0.794	0.850	2.303	0.923

lower, which is a desirable result, so the results presented in the chapter will be for the reduced data set.

With the reduced data set, the benefit of the amount of samples used is looked at in Table 4.3. The best performance occurred with zero or one sample with one Diffs and splines included in the model. In the reduced data set cases (all combinations of Diffs and splines) evaluated with the linear regression there is less than five days difference in the median error between no samples and five samples. The numerical difference in performance as the number of samples increased decreased in some of the methods, but the only case which exhibited a statistical significant performance shift was quadratic with a p-value of 0.05 from 1 to 3 samples and smaller after that.

The behavior of the regression does vary in performance based on the inclusion of the behavioral representation approaches not just the number of samples included. This is seen in Fig. 4.1 with the best performing evaluation technique (SVR with linear

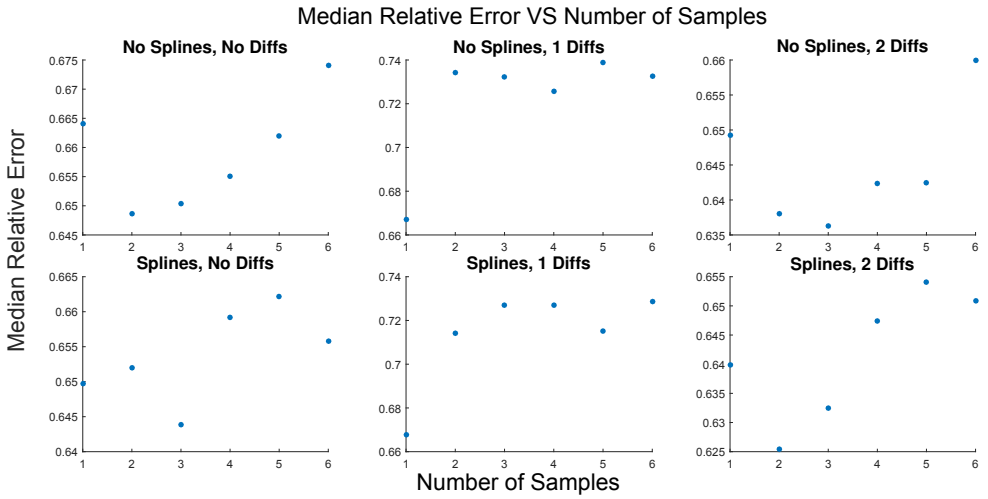


Figure 4.1: SVR linear with reduced data of t_1^* relative error in the different evaluation approaches as a function of the number of samples.

kernel) along with the reiteration of the trend that the median relative error increases with the number of samples included in the evaluation. It is important to note the y-axis here, with the inclusion of the behavioral representation, the median relative error is improved, for example splines and 2 Diffs in the worst combination of samples included performs better then the best of no splines and 1 Diffs.

4.4 Summary

The best performance in this approach is consistent with the observations in the discrete LOS implementation that the performance is improved with the inclusion of behavioral representation while additional samples is not necessarily beneficial, the p-values were greater than 0.1 in most cases. The best models are seen in Table 4.4 all

of which are done with the reduced data set, supporting the reduction in sparseness in the data. One exception is in quadratic, the model with no splines and no Diffs showed a statistically significant improvement to the model with 2 Diffs and splines with a p-value of 0.014.

There is no comparative analysis made to work in the literature due to the uniqueness of the data set and the focus being on the representation with the impact determined by the length of survival regression analysis.

Table 4.4
BEST PERFORMING REGRESSION MODELS USING REDUCED DATA†.

Evaluation Method	**Features Included**	**Median Absolute Error (Days)**	**Median Relative Error**
Linear Reg.	3 Samples, 7 day, 2 Diffs, Splines	51.19	0.765
Quad Reg.	1 Sample 14 day, No Diffs, No Splines	53.05	0.800
GP	0 Samples, 2 Diffs, Splines	52.81	0.822
SVR-Linear	1 Sample, 7 day, 2 Diffs, Splines	32.48	0.629
SVR-RBF	0 Samples, 2 Diffs, Splines	31.48	0.640
Linear Reg.	1 Sample, 14 day, No Diffs, Splines	50.27	0.800
Quad Reg.	0 Samples, No Diffs, Splines	56.05	0.889
GP	1 Sample, 7 Day, No Diffs, Splines	53.37	0.852
SVR-Linear	1 Sample, 1 Diffs, No Splines	31.35	0.619
SVR- RBF	0 Samples, 1 Diffs, Splines	41.60	0.752

†Above the triple line is t_1^* and below is t_2^*.

Chapter 5

Piecewise Aggregate Approximation Fit

The dimensionality reduction method of piecewise aggregate approximation (PAA) is used to represent the time series modes as the input for regression approach of LOS determination. The background was introduced in Chapter 2.1.4. In the PAA approach, the considerations to the multi-modal and nonuniform nature will be addressed. Multiple ways of describing the fit were deliberated with methods taking into account the error in each segment and with respect to the number of samples in each segment.

The analysis of the performance of the standard PAA was then implemented with a

10-fold cross validation approach. The width of the segments for the time series were allowed to be determined for each time series as independent parameter; meaning ALB, HGB, and WT durations were allowed to have independent control of the segment width as if they were single mode observations, like PAA is intended. The starting point of t^* was still used for LOS consistency.

Portage, a high performance computing cluster at Michigan Technological University, was used in obtaining results presented in this chapter.

5.1 PAA Literature Review

In the literature, PAA has been used extensively with similarity searches, clustering, indexing, and queries [15, 17, 31]. The descriptions on fit behaviors have come in to play with the comparison of a series in queries; these searches are of similar length or when different, the approaches of time warping can be used [17]. Alternatively, truncation or zero-padding can be used [31]. There is nothing addressing the fit of the representation though, especially for non-uniform series.

The extensions of data adaptive methods like SAX [134] consider fit focusing on the similarity and queries based methods, not how well the data is being represented to be used in additional analysis.

5.2 Base Adaption of PAA

In expanding the representation, the first phase is to implement a version of PAA as close as possible to the standard form. To implement PAA with the data set used in this research there are several considerations that need to be made and adaptions as a result. To adapt PAA the following issues must be considered:

1. Multiple modes per sample
2. Non-uniform number of observations
3. Non-uniform distribution of the observations
4. Between samples the number of observations vary

Traditionally, PAA takes the time series $X =< x_1, ..., x_N >$ and then creates a new time series to represent it with reduced dimensionality of size n. In this approach each segment is uniform. The creation of each new segment, $\bar{x_i}$ takes the form [31]:

$$\bar{x_i} = \frac{n}{N} \sum_{j=\frac{N}{n}(i-1)+1}^{\frac{N}{n}i} x_j, \tag{5.1}$$

where i is the ith segment in the new series being created and j is the index for the observations in the original time series X with the starting point from where the previous segment ended.

To address issue 1 relating to the data set, each of the modes will be treated independently in creating the reduced dimensionality representation. As approached in [31], when doing similarity searches of different lengths padding sequences of shorter lengths with zeros is an accepted practice, as such this will be used as a starting point. The zero padding is visualized in Fig. 5.1. It is expected that this is not optimal due to the complexity of the nonuniform durations of observations. For the data set there is a maximum of two years of observations, so to adapt PAA to non-uniform lengths I will consider a uniform time of the maximum duration of the full two years of observation. Since the setting is non-uniform in the observational frequency as presented in issue 3 the duration of time between the first and last observation for each sample will be considered. Future extensions to consider the different durations of the observations for the different patients and their respective clinical observations could include using the mean duration of observation.

5.3 Experimental Design

As traditional PAA uses uniform segment sizes, this implementation will consider 4 different quantities of segment windows considered: 4, 8, 12, and 24. The duration of the segment window is two years divided by the maximum number of windows considered. For example, 12 segment windows has a duration of 60 days per window. Standardization will be applied as seen in Fig. 5.2. From this base number of window

segments, the number of windows included will be varied. For example, if there were four windows, there would be four models evaluated, one with only the most recent window, one with half the possible windows, one with three of the four windows and finally with all four windows included. The best number of windows was the window combination that performed the best.

Initial evaluation of PAA was done with the complete patient data set with the exception of one patient who was removed due to large errors in their clinical observations. The data was also standardized.

The regression will be done with linear and quadratic models along with Gaussian Processes having constant basis and a SVR with a linear kernel. A 10 fold cross validation approach will be used, with nested cross validation used to select the cost parameter for the linear kernel with options of $C = \{0.1, 1, 10, 50, 100, 500\}$.

The metrics for evaluation are absolute and relative error looked at from both the mean and median perspectives. The t-test was used to compare against other methods for statistical significance.

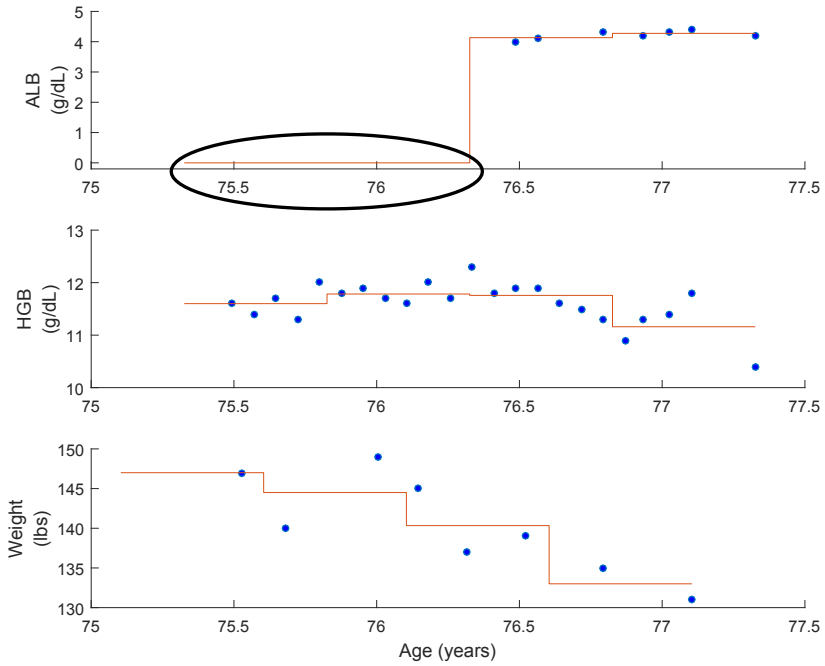


Figure 5.1: The regions of the ALB circled demonstrates zero padding, that the value zero is assigned to the region since no observations exist.

5.4 Results

In the linear regression model, Table 5.1, the best models with t_1^* were consistently occurring with only one segment included, in fact many of the models which performed the best had only 1 or 2 window segments included with only a few exceptions.

Of the models evaluated, the quadratic regression had the highest error. In Table 5.2, the error occurring extends beyond the numerical behavior to the standard deviation occurring with the absolute error. These values are higher then the other methods. With only one exception, the quadratic regression does perform best with only one

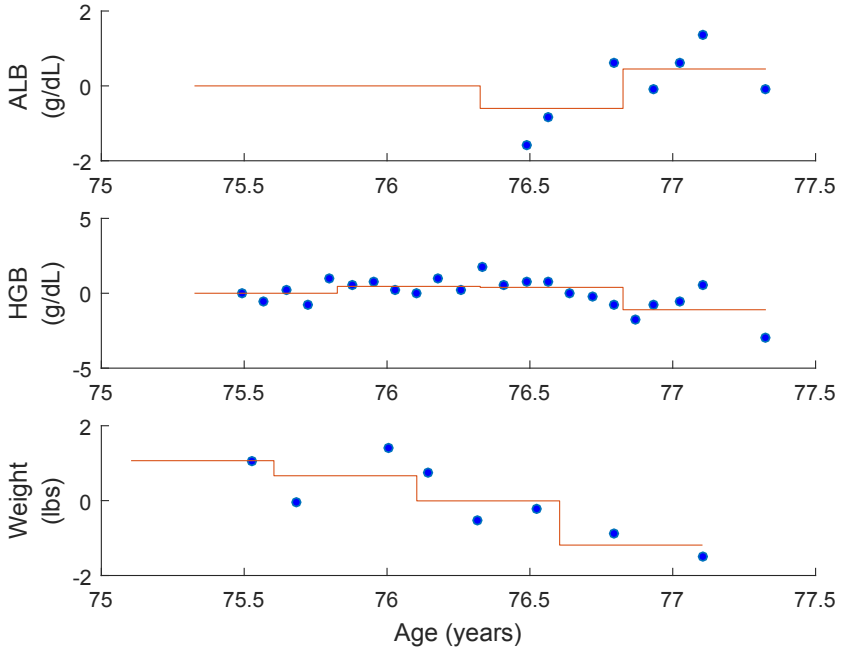


Figure 5.2: PAA applied to the clinical data from Figure 5.1 that has been standardized first. Standardization allows for a the multiple types of clinical data to become comparable in comparing fit error.

Table 5.1
PAA EVALUATED WITH LINEAR REGRESSION USING t_1^* ABOVE THE TRIPLE LINE AND t_2^* BELOW.

Base # of Windows	# Best Windows	Absolute Error (Days)	Median Relative Error
4	1	73.11±84.3	0.798
8	1	73.00±84.5	0.792
12	1	72.75±84.2	0.785
24	1	72.31±84.0	0.787
4	2	73.42±84.8	0.900
8	6	74.19±84.4	0.853
12	5	72.97±85.1	0.844
24	1	72.39±84.2	0.868

window considered.

The GP evaluation also had a similar level of error to the linear regression evaluation

Table 5.2
PAA evaluated with quadratic regression using t_1^* above the triple line and t_2^* below.

Base # of Windows	# Best Windows	Absolute Error (Days)	Median Relative Error
4	1	76.09±86.3	0.839
8	1	77.30±87.1	0.896
12	1	76.75±86.0	0.841
24	2	79.19±86.1	0.887
4	1	76.87±87.8	0.941
8	1	76.99±89.1	0.931
12	1	76.35±88.3	0.916
24	1	76.95±85.5	0.934

approach, see Table 5.3, but was not as high and did not have as much variance as the quadratic repression.

Table 5.3
PAA evaluated with Gaussian Process regression using t_1^* above the triple line and t_2^* below.

Base # of Windows	# Best Windows	Absolute Error	Median Relative Error
4	1	73.64±84.2	0.843
8	1	73.68±84.1	0.847
12	1	73.30±83.8	0.845
24	2	73.13±83.6	0.836
4	2	73.68±84.8	0.937
8	1	73.32±84.4	0.908
12	2	73.28±84.1	0.902
24	2	73.02±83.9	0.894

The last of the methods considered was the SVR with the linear kernel, which resulted in lower median relative error and absolute error but higher standard deviations as seen in Table 5.4. The number of windows included out of the number available continued to be a small fraction.

Table 5.4
PAA evaluated with SVR with a linear kernel using t_1^* above the triple line and t_2^* below.

Base # of Windows	**# Best Windows**	**Absolute Error**	**Median Relative Error**
4	3	64.85±98.3	0.656
8	1	64.59±98.5	0.659
12	1	64.13±98.3	0.648
24	4	63.99±98.0	0.639
4	2	64.47±99.4	0.651
8	1	64.18±99.5	0.644
12	5	64.19±98.8	0.657
24	2	64.23±98.6	0.667

From the number of windows included to get the best performance in the analysis, the sparseness of the matrices and the fit of the PAA was considered. In Fig. 5.3, it is clear that when there is a base of 24 windows, the fraction of the inputs that contain non-zero information is higher than with less windows considered. The 18th window the fraction of input information falls to 50% for both t_1^* and t_2^*.

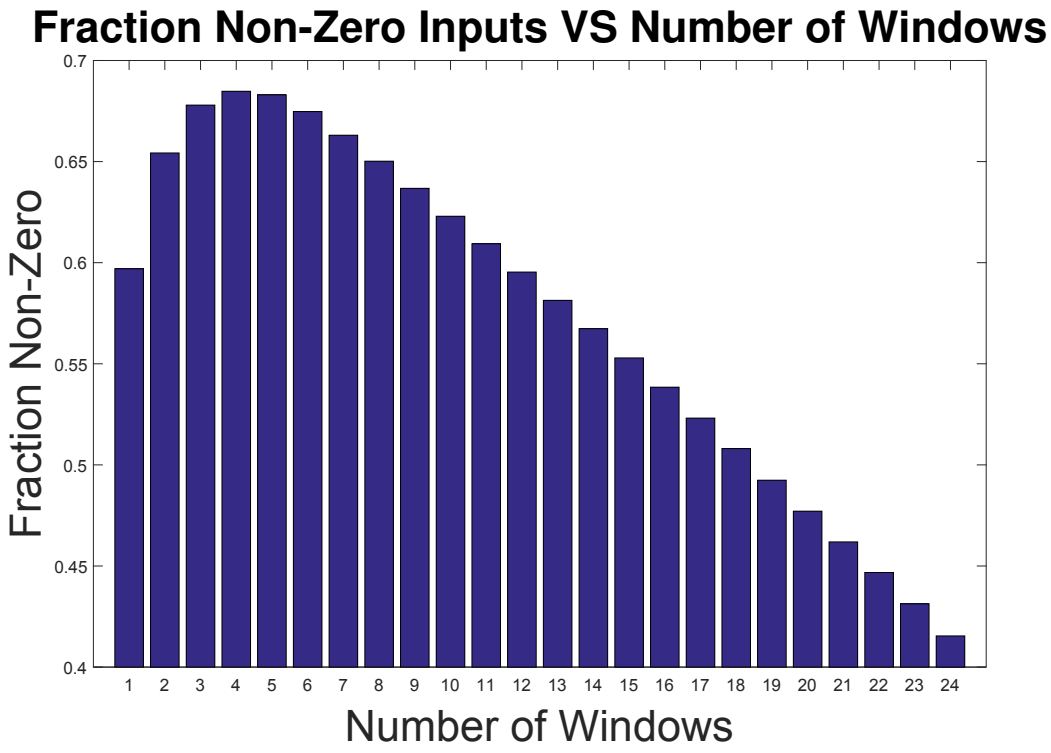


Figure 5.3: The sparseness of the PAA data representation through the fraction of non-zero inputs for the patients with a base of 24 window segments as a function of the number of windows included in the analysis.

5.4.1 Reduced Data Set

As previously applied with the BN evaluation approach, data set reduction was considered (see Section 3.1.2.1). The sparseness factor was observed, and it was determined relevant to apply reduction to this approach. The methodology which was previously established removed all patients which were missing one or more of the clinical observations. This same reduction approach will be used as a pre-processing step to reduce the sparseness of the PAA.

After reducing the data the attention the number of windows considered was looked at again. The number of window segments included out of the possible available does has an effect on performance. In Fig. 5.4, the linear regression, GP, and SVR evaluation methodologies were considered through their median relative errors against the number of windows considered in the evaluation model. With the blue plot representing the GP, there is stability observed, however the linear and SVR exhibit upward trends that as more windows are added into the model the error is increasing.

The results of the evaluations compared with the best regression models using the clinical samples, difference trends and splines are shown in Table 5.5. In this comparison, the PAA method was not numerically an improvement over the previous approach. While the values are not better, there is not a statistical difference for

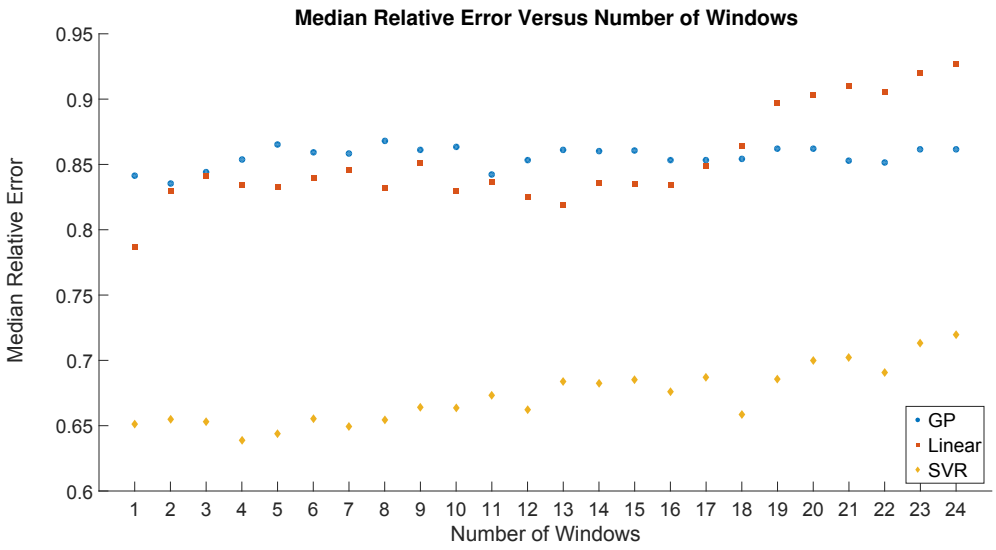


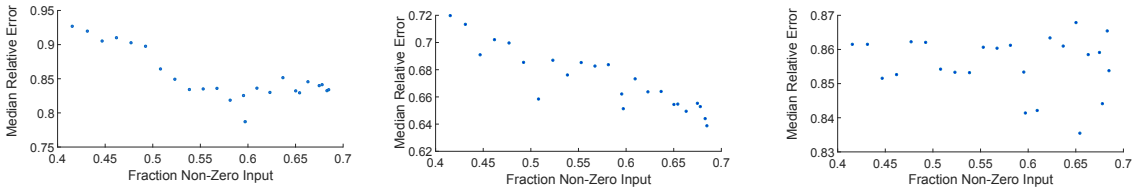
Figure 5.4: The median relative error versus the number of windows for GP (circle marker), linear regression (square marker), and SVR (diamond marker) evaluation methods.

most window segment considerations. The p-values reach as high as 0.66 and as low as 0.12 when the GP evaluation compares the best regression against the possible representations for the window inclusions. The SVR with linear kernel was a more diverse case with some windows of p-values <0.001 and others as high as 0.97.

Table 5.5
BEST PERFORMING REGRESSION MODELS USING REDUCED DATA SET†.

Evaluation Method	**Regression Features Included**	**Median Relative Error**	
		Regression	**PAA**
Linear Regression	3 Samples, 7 day, 2 Diffs, Splines	0.765	0.785
Quad Regression	1 Sample 14 day, No Diffs, No Splines	0.800	0.839
Gaussian Proc.	0 Samples, 2 Diffs, Splines	0.822	0.836
SVR- Linear	1 Sample, 7 day, 2 Diffs, Splines	0.629	0.639
Linear Regression	1 Sample, 14 day, No Diffs, Splines	0.800	0.844
Quad Regression	0 Samples, No Diffs, Splines	0.889	0.916
Gaussian Proc.	1 Sample, 7 Day, No Diffs, Splines	0.852	0.894
SVR- Linear	1 Sample, 1 Diffs, No Splines	0.619	0.644

†Above the triple line is t_1^* and blow is t_2^*.



(a) Linear regression (b) SVR (c) GP

Figure 5.5: Median relative error versus the amount of non-zero inputs to the model for 1-24 windows.

One of the relationships with this representation and the evaluation performance is the sparseness, which is interpreted through what fraction of the input segments are non-zero. This relationship is visualized in the Fig. 5.5. With the relations between the fraction non-zero and the performance, the linear (Fig. 5.5(a)) and SVR (Fig. 5.5(b)) display the clearest correlations that as the fraction of non-zero inputs increases, meaning more of the inputs contain values, the median relative error decreases in a trend approaching a linear behavior.

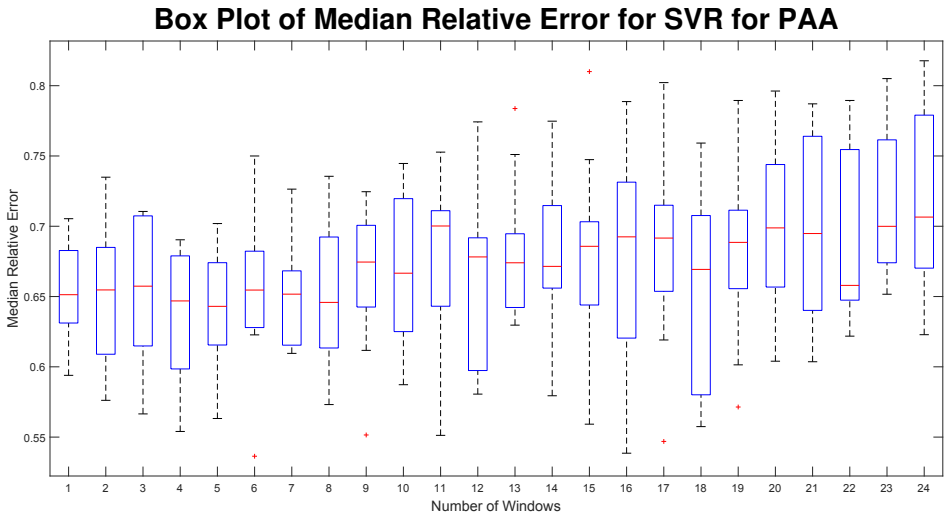


Figure 5.6: Boxplot of the median relative error for the SVR evaluation with 1-24 windows.

The box plots for the evaluations are representative of how the median relative error for each fold is distributed, shown in Figs. 5.6 - 5.8. This is an important performance factor since there is a large distribution in error occurring. One of the primary characteristics seen with the GP in Fig. 5.8 is the stability that occurs starting around the 12th window that the data is not changing any further when more of the segments are considered do not contain much more clinical information to effect the performance. The other methods are not as stable, they in turn displays the broader impact from the higher level of sparseness occurring in the representation with the higher window count.

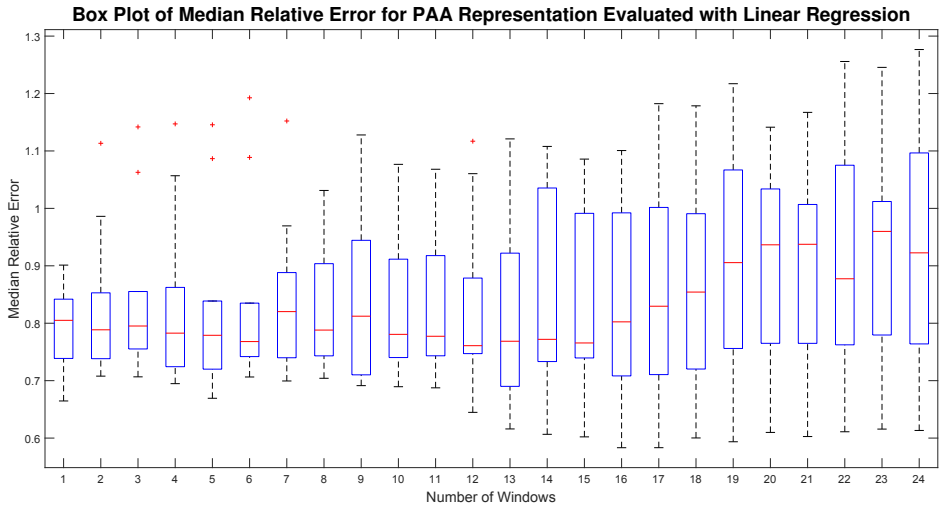


Figure 5.7: Boxplot of the median relative error for the linear regression evaluation with 1-24 windows.

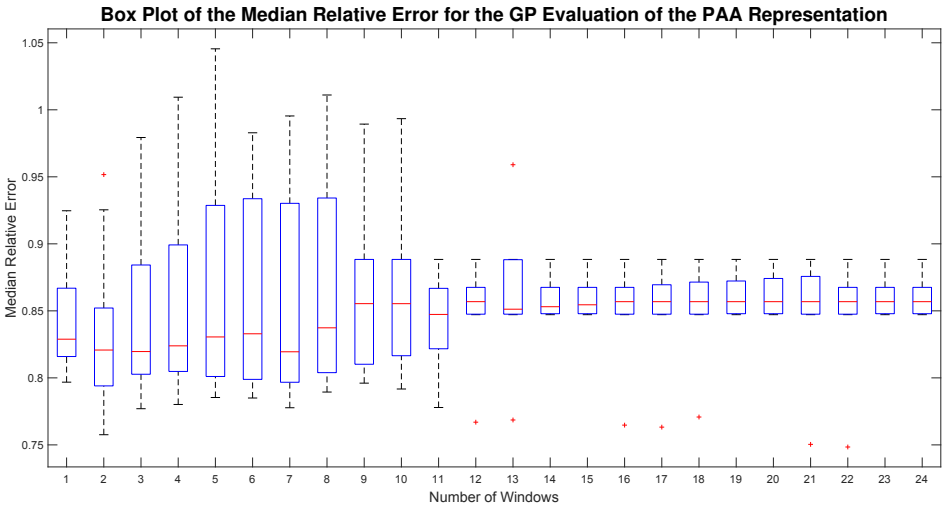


Figure 5.8: Boxplot of the median relative error for the GP regression evaluation with 1-24 windows.

One of the issues encountered was the high error in the quadratic regression. This was occurring due to rank deficit during the evaluation process, that there wasn't sufficient differential in the values to create a fit appropriate. Part of the reason for this behavior is explained with the fraction non-zero input. After the amount of information that is not zero drops below about 65% it becomes unable to fit the data and handle all the zeros. One of the solutions looked at was stepwise quadratic regression, which drops terms of the quadratic function during the fit if they are unable to be determined rather then letting the rank deficit take over the model. This approach did help some in the 4 and 8 segment versions, in the 12 segment version it took a month to evaluate and after the inclusion of the 8th window the performance began to deteriorate again. Even in the best cases it did not approach the performance of the SVR and with the amount of time to evaluate, this is an infeasible evaluation method.

5.5 Summary

There are two key take aways: impact the expansion of data adaptive segmentation methodology. The first is the high degree of sparseness with the 24 window versions, this warrants no need for more then 12 windows to be considered. The other consideration is the high variability and substantially worse performance of the quadratic regression compared to the other methodologies. Overall the idea of more data does

not mean better performance reappears as it did in previous chapters, this is supported with Table 5.6, that only a few of the windows available need to be included to obtain the best possible results.

Table 5.6
BEST PERFORMING MODELS. ALL THE BEST MODELS WERE WITH t_1^*.

Method	**Num of Win /Base Wins**	**Absolute Error**	**Median Relative Error**
Linear	1/12	72.75±84.2	0.7854
Quadratic	1/4	76.09±86.3	0.8388
GP	2/24	79.19±86.1	0.8355
SVR	4/24	63.99±98.0	0.6388

While the results of the PAA were not an improvement upon the regression, there is not a statistical difference, with the exception of the quadratic regression case. This method has the benefit of dimensionality reduction simpler then the clinical sample extraction, difference trends and splines used with the previous regression methodology.

There is no comparative analysis made to work in the literature due to the uniqueness of the data set and the focus being on the representation with the impact determined by the length of survival regression analysis. Literature focus with this representation approach is on similarity search and indexing, which is not comparable.

Chapter 6

Piecewise Aggregate Approximation Segmentation

The adaptation of piecewise aggregate approximation (PAA) to the data using the fit is the challenge that is approached based on the fit parameter developed. This is unique based on the previous data adaptive modifications of PAA being focused on the magnitude of the data post segmentation. Consideration of segmentation approaches will also be established but face difficulty in the nonuniform multi-modal nature of the data set being used.

Portage, a high performance computing cluster at Michigan Technological University, was used in obtaining results presented in this chapter.

6.1 Data Adaptive PAA

To extend PAA, adapting to the data is the logical approach. In the literature this has been approached through the modification of the breakpoints of symbolic aggregate approximation (SAX) using differential evolution and genetic algorithms [18, 54]. Another approach has been to adapt the representation to contain more information, for example instead of just the SAX symbolic assignment three values: min, max and aggregate values can be used [14].

The adaptive piecewise constant approximation (APCA) at first appearance seems to give the same result of an adaptive PAA. This approach depends on the time series undergoing the Haar Discrete Wavelet Transform to get the coefficients for the segment boundaries [135]. This is not something that is challenged when applied to non-uniform time series just as other transforms encounter like the DFT. It also does not natively have the ability to manage multiple time series of information when considering the segment boundary placements. The solution to the distribution problem that was presented was to use minimum bounding rectangles to represent regions of data using rectangle feature points rather then all the points in the segment [135].

6.2 Segmentation of Time Series

Other considerations can be made for window segmentation approaches that are done to time series since that is a similar problem.

To start segmentation is defined as the problem given a time series T produce the best representation with K segments that each of the segments does not exceed the maximum error specified and neither does the combined error across all the segments. Traditionally done with one of three approaches: top down, bottom up or sliding window [136]. It can also be done online as the data is collected or offline (or in a batch format with all the data collected) [137].

If the problem of segmentation has known segmentation results, the Beeferman segmentation metric is what would be used to describe the accuracy of the segment placement [138]. This metric is based on the error between the placement of the observation in the segmented series and where it should be placed based on the true segmentation. If the segmentation is not known, least squares is a common error metric used [137]. Segmentation also suffers from the problem of loosing information about behavioral changes, that it is possible for the behaviors (such as if the slope is increasing or decreasing) to change multiple times in a single segment depending on the placement [139]. This issue lead to the development of a clustering based

constraint approach that limits the amount of behavior changes in in a segment [140].

An evolutionary approach using genetic algorithms allows for the chromosomes to be comprised of the segment endpoints to recognize the significant points, recognizing that significant points are lost by using uniform segment widths [141]. This approach has also makes use of a distance metric that is slightly modified to consider the distance between significant points via their amplitudes and temporal distances in the fitness function. With the lack of homogeneity across data makes segmentation difficult, another approach that is possible is to represent the chromosomes as a floating point value [16] and combine this with a rule based system to aim for homogeneous behavior within segments.

Segmentation can also be augmented with information to tell how much data was in an region prior to dimensionality reduction being applied. One case was when applying APCA to describe how much data was in each segment [50].

6.3 Methodology

To create an adaptive segmentation, the basis is optimization, this will be done through the utilization of genetic algorithms (GA). Different from the traditional segmentation problem there are three time series that are considered at one time

instead of one. Also different, instead of a user specified error condition the aim is to use an optimization function which initially combines the fit error and the performance initially, and later on the fraction of non-zero window segments. With the optimization a function is necessary, this is based on the fit error. Fit error as shown in the previous chapter is complicated by the sparseness and nonuniformity of the time series data.

Given there are three time series for each patient, the segmentation boundaries need to occur at the same locations across all series. With the varying lengths of the data observations, zero padding of the shorter series will be done but the lengths of the shorter series will need to be considered as options for where to place the segments.

Evaluation will be conducted with linear regression, Gaussian Process with a constant basis, and support vector regression (SVR) with a linear regression. The cross validation, parameters, and standardization for the window segments are the same approach which were used with the standard PAA (see Chapter 5).

6.3.1 Optimization

A standard formation of the optimization function is comprised of a loss component and a cost component with a weight parameter λ. The loss feature represents the

incorporation of the error from the evaluation and the cost is from the fit of the PAA,

$$f = loss + \lambda cost. \tag{6.1}$$

This optimization function process is visualized in Fig. 6.1. For the cost parameter in the optimization function, the basis is the Euclidean Distance with a consideration on the number of observations. Part of the reasoning for the decision for the use of Euclidean Distance is the commonality of the approach. Typical distance metric used in the comparison of series include the magnitude of the difference or the square of the difference [142]. Euclidean Distance is an extension on the squared approach taking the square root following the square.

For the optimization, the function is based on the euclidean distance of the fit of the PAA for the segments to the points that exist for each of the three time series.

The distance, PAA_d, is calculated based on the summation of Euclidean distances in each segment for each point,

$$PAA_d = \sum_{i=1}^{seg} \frac{\sqrt{\sum_{j=1}^{k_i}(PAA_i - x_j)^2}}{k_i}, \tag{6.2}$$

where, seg is the number of segments, and k_i is the number of samples in segment i. This considers the nonuniformity of the samples, since the error distance in each

segment is then divided by the number of clinical samples in that segment for the patient. This helps avoid the bias from forming if the PAA is fit to one sample or fifty.

Since the objective is to determine the LOS performance, the relative error will play a role in the optimization. To avoid the different durations playing a large role in the performance, the maximum duration observation of two years will be used and zero padding will be used to account for when it is beyond the observation period. As it was observed in the PAA exploration, the zero padding does have a substantial negative impact on the relative and absolute error when predominate, to avoid this impact the performance is considered in two components for the loss term: one segment and half segments considered in the evaluation.

The final evaluation on the test set once the segmentation is determined will considered all possible segment inclusions from 1 segment through to the maximum number possible. This is done as a consideration to the duration which should be considered and the impact of zero padding in the data adaptive approach.

The general approach is seen in Fig. 6.1 and an expanded representation of the fitness function used in this first approach is seen in Appendix D.2.

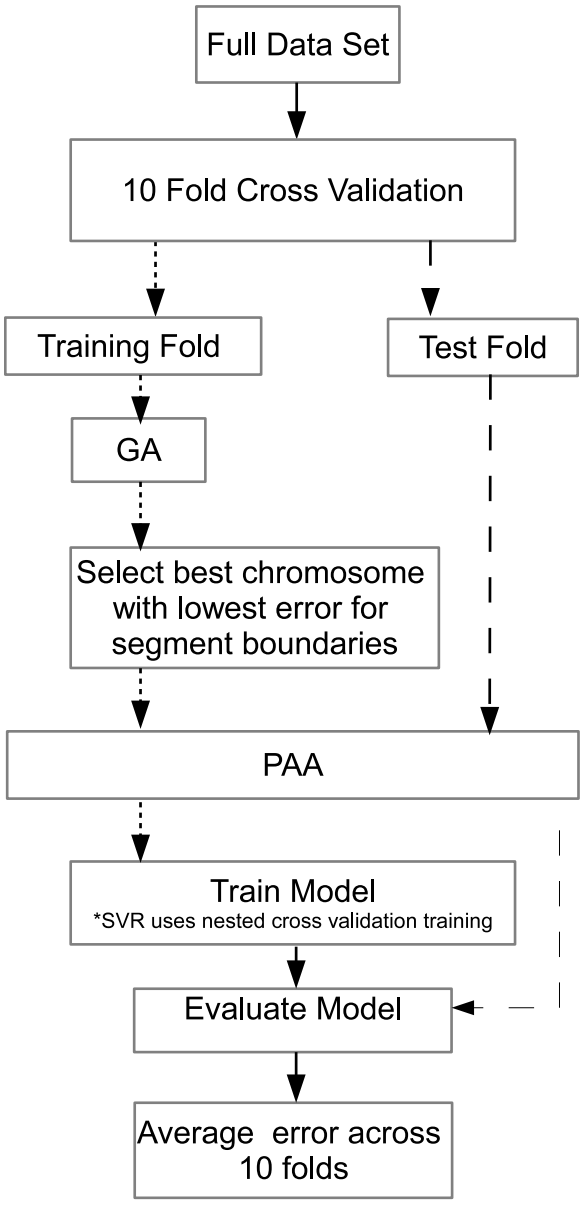


Figure 6.1: Validation methodology for PAA-GA.

6.3.2 GA Parameters

The chromosomes represent the number of segments -1 split points as a vector. With the number of window segments equal to 4, 8, and 12. The 24 window version was

not done in this experiment due to the performance observed in the standard PAA application (see Chapter 5).

The other parameters to consider for the GA include the crossover rate $\{0.3, 0.5, 0.7\}$, crossover ratio $\{1.2, 1.4, 1.6\}$, and the mutation rate $\{0.2, 0.4, 0.6\}$. Population size of 50 with a maximum of 100 generations was used. A minimum window size of 7 days was used.

6.3.3 Evaluation of Initial Fitness Function Version

During the evaluation with the linear regression there was rank deficit warning occurring, this is to be expected. With the large number of zeros present to start and the zero padding this occurs with the larger number of segments. There is a known bias [143] with zero padding that occurs. For this reason the fitness function considered the single segment evaluation, which did not experience rank deficit warnings in evaluation.

During the evaluation the computational time for the SVR approach is unfeasible due to the high volume of nested cross validation within each of the GA individuals and each iteration.

6.4 A Practical Optimization Function

While the theory to include the performance is sound, it is not practical. The results from the GP and linear regression show promise from a factor in the consideration of the performance still. Based on the non-adaptive approach there is a fairly clear correlation with the median relative error and the fraction of the non-zero inputs to the model, which was seen in Fig. 5.5(b) for the SVR to have nearly a linear trend with a couple outliers where the relative error was lower sooner.

The linear regression leveled off after a while in the observation. The worst performing of the standard PAA approach also exhibited the behavior that the higher the level of non-zero input the better performance, be it after a cut off. The GP was the only approach to not support this logic directly with a more uniform behavior. The best performing model with the PAA was the least successful in the first fitness function model that was used due to the computationally intensive nature.

Due to the reasonable computation time to compute the PAA, the fraction non-zero input will replace the performance factor, creating the method PAA-GA. This produces the fitness function,

$$f = d + \lambda(nonzero_1 + nonzero_2). \tag{6.3}$$

There are two non-zero factors, one ($nonzero_1$) to incorporate the sparseness for one of the windows and the other ($nonzero_2$) for half of the available windows. This was done since the number of windows included in the PAA experiment effected performance and had varying levels of sparseness.

6.5 Results

The incorporation of the fitness function based on the fraction non-zero representation is computationally more efficient then the first approach. It was a noticeable observation of days to weeks less to get results of the SVR computation back compared to the original fitness function.

To determine the worth of the representation, it is compared against the representation without the GA to optimize the placement. The comparison was done with the 12 window option since that was the case which observed the best performing case in the standard PAA. The performance analysis was considered based on the GA parameters (crossover rate, crossover ratio, and mutation rate) which produced the best accuracy for the final evaluation. The best example of the compared performance was observed in comparing with the standard PAA with the Gaussian Process (GP) evaluation seen in Fig. 6.2. The PAA-GA method consistently had lower error rates then the standard PAA.

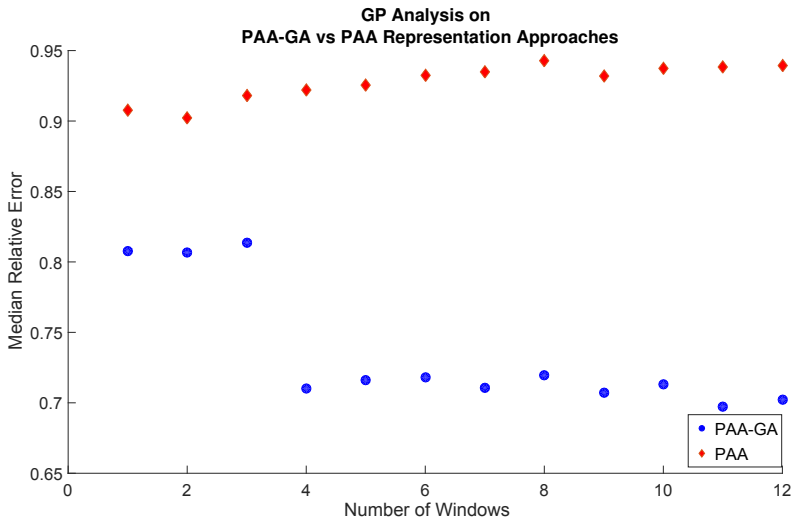


Figure 6.2: Best GP with GA for segment placement compared with the PAA. Mutation rate of 0.6 with a crossover rate of 0.3 and crossover ratio of 1.6, note the blue points are the PAA-GA method and the red diamonds are the standard PAA.

Evaluating with SVR and a linear kernel was not as clean in the performance benefits. Fig. 6.3 shows that for 2-4 windows included the difference in error between the methods was not only not substantially different, but the performance was superior in the standard cases. The GA parameters which produced the best accuracy here were not the same as that of the GP best performance, this shows that there is a fit factor and other parameters may lead to full improved performance in all cases. In the PAA methodology the performance diminished with the inclusion of more windows, the PAA-GA method mediated this issue with improved performance when more windows are included.

The linear regression evaluation displayed a unique feature different from the other two evaluations, that the higher number of windows did not always out perform,

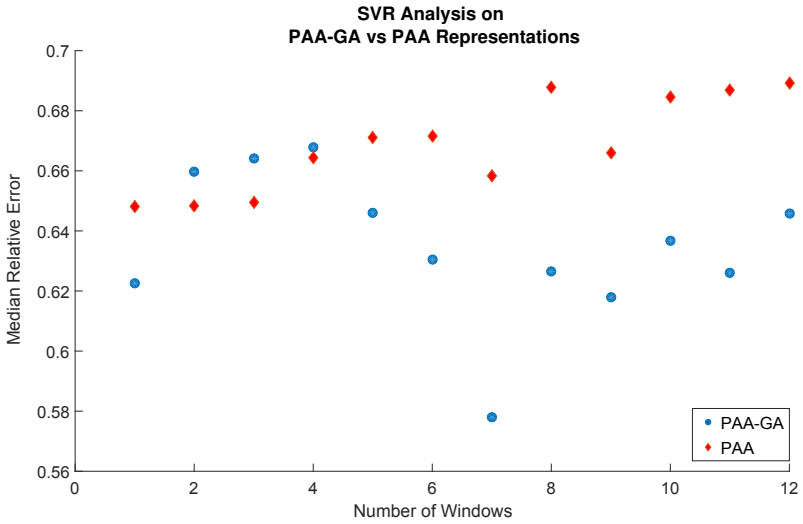


Figure 6.3: Best SVR with GA for segment placement compared with the PAA. Mutation rate of 0.2 or 0.4 with a crossover rate of 0.5 and crossover ratio of 1.2, note the blue points are the PAA-GA method and the red diamonds are the standard PAA.

that that was the region where performance benefits were of question. In Fig. 6.4 the PAA-GA method improves upon the accuracy except for when there are 9-11 windows, in those cases it is insignificantly close or performs worse.

Overall, the performance is an improvement compared to the previous method. One of the noted issues previously was outliers and the high variance levels. To characterize this behavior, boxplots were used for the best performing cases. In the linear regression case, Fig. 6.5, less then half of the window inclusion options encountered cases which outliers occurred. The median relative error across the folds in all cases is maintained below 100% error. The third quartile error in come cases does extend beyond 100% median relative error on the folds, this is in part a product of the LOS distributions which some are a few days and some are hundreds making the regressive

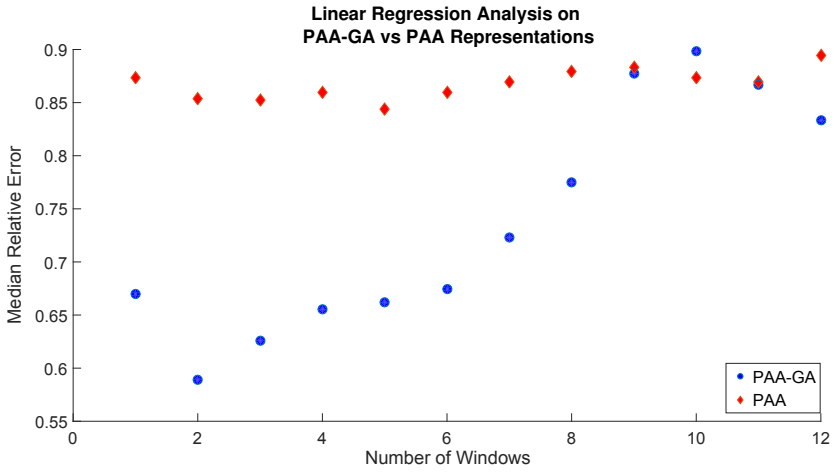


Figure 6.4: Best linear regression with GA for segment placement compared with the PAA. Mutation rate of 0.2 with a crossover rate of 0.5 and crossover ratio of 1.6, note the blue points are the PAA-GA method and the red diamonds are the standard PAA.

fit more difficult.

The SVR analysis showed the greatest stability with no outliers occurring, Fig. 6.6.The variation in the cap of the third quartile was fairly unremarkable in hat it was consistent around 0.75% median relative error.

On the other end of the spectrum, the GP evaluation consistently produced outliers, Fig. 6.7. With the exception of the first three window versions the maximum on the range of the median relative errors for the folds stays below 100%.

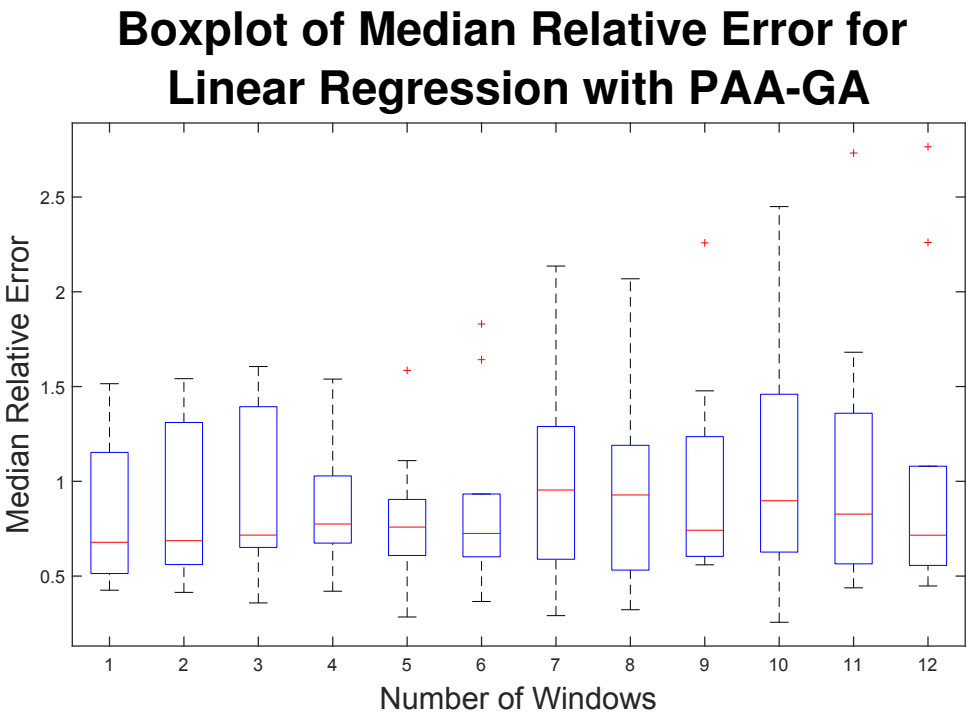


Figure 6.5: Boxplot of the ten folds for the Linear regression analysis of the best GA parameters, median relative error for 1-12 windows. Mutation rate of 0.2 with a crossover rate of 0.5 and crossover ratio of 1.6

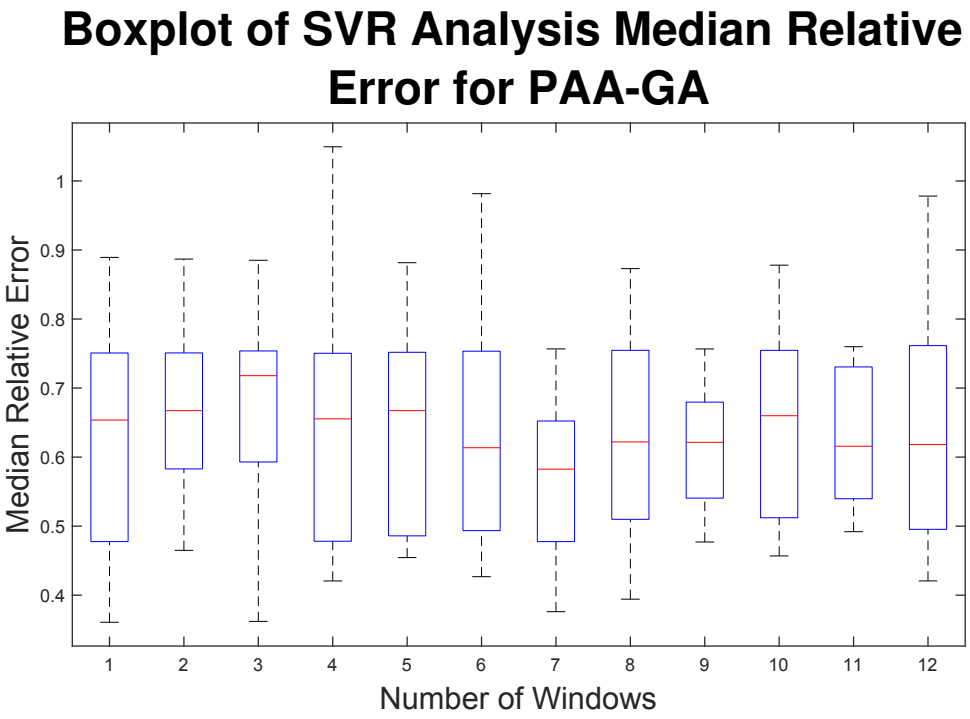


Figure 6.6: Boxplot of the ten folds for the SVR regression analysis of the best GA parameters, median relative error for 1-12 windows. Mutation rate of 0.2 with a crossover rate of 0.5 and crossover ratio of 1.2

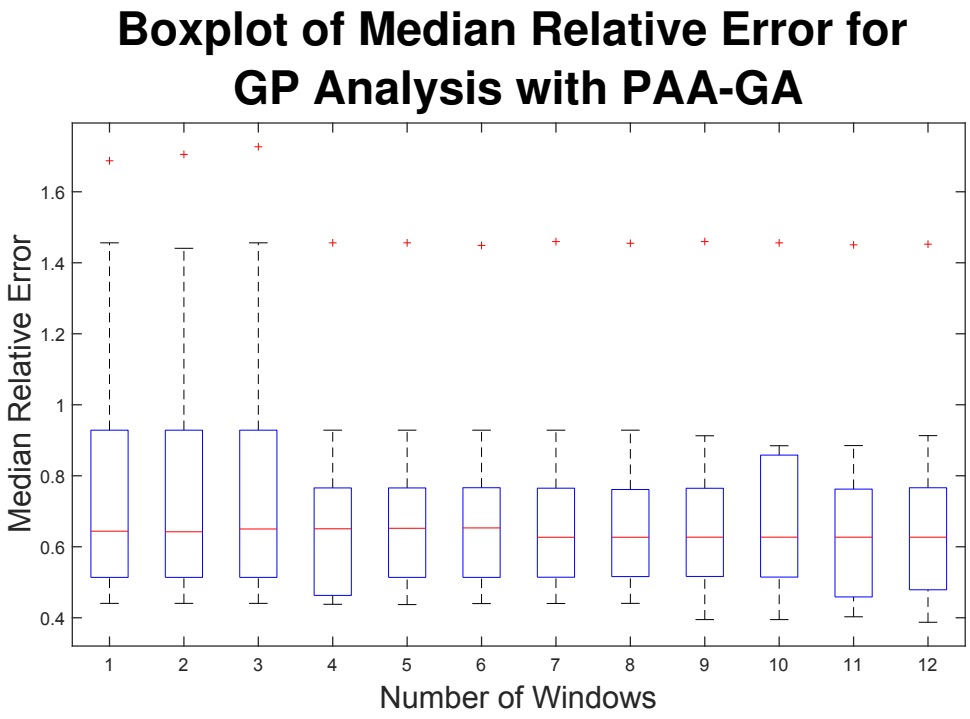


Figure 6.7: Boxplot of the ten folds for the GP regression analysis of the best GA parameters, median relative error for 1-12 windows. Mutation rate of 0.6 with a crossover rate of 0.3 and crossover ratio of 1.6

6.6 Summary

The incorporation of the genetic algorithm has helped improve how the regression analysis performs overall. Depending on the fitness function the time to determine the segmentation is computationally infeasible. From the results with the range of GA parameters considered there was improvement of 20% in some scenarios.

While there was a significant in some cases improvement in the overall performance, there was not any statistical improvement seen, this was done with a t-test on the best GA parameter set. For the linear regression, the smallest p-value from the t-test was 0.326 with one segment window. The SVR method was expected to not see cases of statistical difference since the curves in Fig. 6.3 were so close, but with 11 windows a p-value of 0.085 was achieved, while not significant, it is promising. GP had several p-values which were around 0.1, with the most significant being 0.076 with 11 segment windows. The reason for this is supported with the box plots, while the median relative error is lower there is a broad range of values in both the original PAA and the PAA-GA which go into creating that value, and that distribution when comparing them reduces the significance of the improved performance.

Due to the standard PAA representation not being an improvement over the clinical sample and behavioral representations previously considered, this was also compared

for t_1^* in Table 6.1. The PAA-GA method has a lower median relative error compared with the regression results. Statistically due to the distributions there is not a difference between the regression and the PAA-GA, with p-values ranging from 0.232 to 0.340 when compared using a t-test.

There is no comparative analysis made to work in the literature due to the uniqueness of the data set and the focus being on the representation with the impact determined by the length of survival regression analysis. Similar to the constraints of the PAA, the work using this method focuses on similarity search and indexing, which is not comparable.

Table 6.1
BEST PERFORMING REGRESSION MODELS USING REDUCED DATA SET WITH t_1^*.

Evaluation Method	**Regression Features Included**	**Median Relative Error**	
		Regression	**PAA-GA**
Linear Regression	3 Samples, 7 day, 2 Diffs, Splines	0.765	0.589
Gaussian Proc.	0 Samples, 2 Diffs, Splines	0.822	0.697
SVR- Linear	1 Sample, 7 day, 2 Diffs, Splines	0.629	0.576

6.7 Future considerations

With the variations shown from the different GA parameters, the parameter consideration is an option which can be further explored. In this exploration, 27 variations of parameters were considered but there are other parameters which can be varied and additional values for the parameters that were considered.

Since there was no significant performance difference between the feedback approach and the non-zero approach it is possible that another aspect of the data or how its represented be considered in the fitness function.

One of the biggest issues observed correlated with sparseness, which is mostly the product of zero padding to account for the variation in the duration. A consideration of truncation in combination to zero padding to reduce the sparseness may improve performance.

One of the parameters in the GA was the fit of the PAA to the data, this can be modified with an alternative approach to Euclidean Distance. Using a minimum bounding distance approach to describe the fit with respect to the dispersion of the points in the segment may be a better approach.

Chapter 7

Conclusion

Work was conducted to determine the length of survival prognosis through various approaches of analysis and representation. The adaptation of the representation techniques to fit the constraints of the unique nonuniform, multi-modal, time-limited time series data was one of the challenges addressed, with more substantial novelty in the adaption of the piecewise aggregate approximation. The genetic algorithm adapted piecewise aggregate approximation method is a newly developed approach to address the specific challenges of the data.

Through the clinical samples and behavioral representation approaches it was established that there was a connection between the representation of the data and the analysis in how performance of length of survival prognosis in both classification and

regression. Adapting the dimensionality reduction technique piecewise aggregate approximation to the nonuniform multi-modal data was not an improvement on the survival performance compared with the clinical samples and behavioral representation previously used. The explanation was associated with the sparseness of the representation, which lead to the development of the piecewise aggregate approximation genetic algorithm method.

With the incorporation of genetic algorithms, there was improvement of up to 20% over the standard piecewise aggregate approximation approach. This method was numerically better then the original behavioral and interpolation approach. The one caveat with the improved survival prognosis is that it is not statistical with the t-test results, explainable with the large distribution in actual survival times that are trying to be modeled.

The future of this work is to consider different modifications of the genetic algorithm parameters. This is important since performance varied with the cross over ratios and mutation rates. Another factor to consider is the observation duration. In the work zero padding was done to extend, fill in additional time segments with zeros, when those times extended beyond the observational time for a given patient. One consideration might be a truncation to a shorter maximum duration to reduce the amount of zero padding needed at the expense of some observations.

This work focused on the application of outpatient oncology, there are other applications as well which present nonuniform, multi-modal, time-limited time series data that this work can be expanded to. Social media data mining is an application which shares many of the same characteristics as the data set which has been used, it has multiple types of data which is observed at various times, such as when users post images versus status versus check-ins. In nature there are observation types which occur irregularly, such as earthquakes and eruptions. In those cases while the sampling devices are set to record at uniform frequencies, the durations of the observation can vary extensively and missing data can occur that can cause issues. Strictly time series or multi-modal sets in other applications are also potentials for future application; such as stock markets, microgrid control, and image analysis done through time series representations.

Even with a piecewise approach interpolation for regions where information is missing is difficult, especially when there is not a uniformly distributed set of samples. The current approaches to interpolating are based on full signal reconstruction or nearest neighbor assignment [26, 27, 35, 144]. It is possible that an entire attribute contains no observations, this eliminates the possibility for interpolation. When data for an entire observation is missing research has traditionally chosen to filter out those samples and thus reducing the sample set size [11, 44, 120, 121]. In addition to the sample size reduction, the behaviors in the other attributes are not considered in the learning process if the sample would be in a training set (cross validation is used to allow

all data to train and test). An alternative to removing the sample is to use a class assignment or assign zero value in place of reduction as a preliminary solution, with other considerations as exploration extensions. This issue extends to normalization and symbolic assignments. They face a challenge due to missing data issues and domain specific population diversity. A traditional zero mean unit variance approach struggles with the aspect of missing data.

Through this work, the challenge of quantifying the amount of data remains. The piecewise aggregate approximation with the genetic algorithms mediates the quantification issue through the number of segments that are zero being the controlling factor.

References

[1] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A symbolic representation of time series, with implications for streaming algorithms," in *Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, DMKD '03, (New York, NY, USA), pp. 2–11, ACM, 2003.

[2] P. Montalto, M. Aliotta, A. Cannata, C. Cassisi, and A. Pulvirenti, *Advances in Data Mining Knowledge Discovery and Applications*. INTECH Open Access Publisher, 2012.

[3] "Support vector machines." Webpage. `http://www.mathworks.com/help/stats/support-vector-machines-svm.html`.

[4] J. Winikus and L. E. Brown, "Representation and incorporation of clinical information in outpatient oncology prognosis using bayesian networks and na ïve bayes," in *2016 IEEE International Conference on Electro Information Technology (EIT)*, pp. 0653–0658, May 2016.

[5] C. Souza, "Kernel functions for machine learning applications," 2010. http://crsouza.com/2010/03/kernel-functions-for-machine-learning-applications/, Accessed: 2015-10-01.

[6] J. Mager, U. Paasche, and B. Sick, "Forecasting financial time series with support vector machines based on dynamic kernels," in *Soft Computing in Industrial Applications, 2008. SMCia '08. IEEE Conference on*, pp. 252–257, June 2008.

[7] C. Yan, J. Fang, L. Wu, and S. Ma, "An approach of time series piecewise linear representation based on local maximum minimum and extremum," *Journal of Information and Computational Science*, vol. 10:0, pp. 2747 – 2756, 2013.

[8] S. Anand, A. Smith, P. Hamilton, J. Anand, J. Hughes, and P. Bartels, "An evaluation of intelligent prognostic systems for colorectal cancer," *Artificial Intelligence in Medicine*, vol. 15, no. 2, pp. 193 – 214, 1999.

[9] J. Hayward, S. A. Alvarez, C. Ruiz, M. Sullivan, J. Tseng, and G. Whaen, "Machine learning of clinical performance in pancreatic cancer database," *Artificial Intelligence in Medicine*, vol. 49, pp. 187–195, 2010.

[10] W. Kim, K. S. Kim, J. E. Lee, D.-Y. Noh, S.-W. Kim, Y. S. Jung, M. Y. Park, and R. W. Park, "Development of novel breast cancer recurrence prediction

model using support vector machine," *Journal of Breast Cancer*, vol. 15, no. 2, pp. 230–238, 2012.

[11] J. M. Jerez, I. Molina, P. J. García-Laencina, E. Alba, N. Ribelles, M. Martín, and L. Franco, "Missing data imputation using statistical and machine learning methods in a real breast cancer problem," *Artificial Intelligence in Medicine*, vol. 50, no. 2, pp. 105 – 115, 2010.

[12] I. Anagnostopoulos, C. Anagnostopoulos, D. Vergados, A. Rouskas, and G. Kormentzas, "The Wisconsin breast cancer problem: Diagnosis and TTR/DFS time prognosis using probabilistic and generalised regression information classifiers.," *Oncology Reports*, pp. 975–981, 2006.

[13] J. Li, D. Maier, K. Tufte, V. Papadimos, and P. A. Tucker, "No pane, no gain: Efficient evaluation of sliding-window aggregates over data streams," *SIGMOD Rec.*, vol. 34, pp. 39–44, Mar. 2005.

[14] B. Lkhagva, Y. Suzuki, and K. Kawagoe, "Extended SAX: Extension of symbolic aggregate approximation for financial time series data representation," *DEWS2006 4A-i8*, 2006.

[15] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra, "Locally adaptive dimensionality reduction for indexing large time series databases," *SIGMOD Rec.*, vol. 30, pp. 151–162, May 2001.

[16] A. Gacek and W. Pedrycz, "A genetic segmentation of ECG signals," *Biomedical Engineering, IEEE Transactions on*, vol. 50, pp. 1203–1208, Oct 2003.

[17] E. J. Keogh and M. J. Pazzani, "Scaling up dynamic time warping for datamining applications," in *In Proc. 6th Int. Conf. on Knowledge Discovery and Data Mining*, pp. 285–289, 2000.

[18] M. M. M. Fuad, "Genetic algorithms-based symbolic aggregate approximation," in *Proceedings of the 14th International Conference on Data Warehousing and Knowledge Discovery*, DaWaK'12, (Berlin, Heidelberg), pp. 105–116, Springer-Verlag, 2012.

[19] J. Lin, E. Keogh, L. Wei, and S. Lonardi, "Experiencing SAX: A novel symbolic representation of time series," *Data Min. Knowl. Discov.*, vol. 15, pp. 107–144, Oct. 2007.

[20] C. Shannon, "A mathematical theory of communication," *Bell System Technical Journal, The*, vol. 27, pp. 379–423, July 1948.

[21] F. Marvasti, *Nonuniform sampling : Theory and practice*. New York: Kluwer Academic/Plenum Publishers, 2001.

[22] H. Haken and J. Portugali, *Information Adaptation: The Interplay Between Shannon Information and Semantic Information in Cognition*. Springer, 2014.

[23] L. Wang, J. Zhan, C. Luo, Y. Zhu, Q. Yang, Y. He, W. Gao, Z. Jia, Y. Shi, S. Zhang, *et al.*, "Bigdatabench: A big data benchmark suite from internet services," in *High Performance Computer Architecture (HPCA), 2014 IEEE 20th International Symposium on*, pp. 488–499, IEEE, 2014.

[24] J. S. Vitter, "External memory algorithms and data structures: Dealing with massive data," *ACM Comput. Surv.*, vol. 33, pp. 209–271, June 2001.

[25] E. Candes and M. Wakin, "An introduction to compressive sampling," *Signal Processing Magazine, IEEE*, vol. 25, pp. 21–30, March 2008.

[26] M. Sacchi, T. Ulrych, and C. Walker, "Interpolation and extrapolation using a high-resolution discrete Fourier transform," *Signal Processing, IEEE Transactions on*, vol. 46, pp. 31–38, Jan 1998.

[27] C. Ford and D. Etter, "Wavelet basis reconstruction of nonuniformly sampled data," *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on*, vol. 45, pp. 1165–1168, Aug 1998.

[28] B. Xingli and Z. Chengjian, "Research on time series forecasting model based on support vector machines," in *Measuring Technology and Mechatronics Automation (ICMTMA), 2010 International Conference on*, vol. 3, pp. 227–230, March 2010.

[29] R. Agrawal, C. Faloutsos, and A. N. Swami, "Efficient similarity search in sequence databases," in *Proceedings of the 4th International Conference on Foundations of Data Organization and Algorithms*, FODO '93, (London, UK, UK), pp. 69–84, Springer-Verlag, 1993.

[30] K.-P. Chan and A.-C. Fu, "Efficient time series matching by wavelets," in *Data Engineering, 1999. Proceedings., 15th International Conference on*, pp. 126–133, Mar 1999.

[31] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra, "Dimensionality reduction for fast similarity search in large time series databases," *Knowledge and Information Systems*, vol. 3, no. 3, pp. 263–286, 2001.

[32] J. Winikus and L. E. Brown, "Representation of Clinical Information in Outpatient Oncology for Prognosis Using Regression," *Advances in Science, Technology and Engineering Systems Journal*, vol. 1, no. 5, pp. 16–20, 2016.

[33] Y. N. Jeng and Y.-C. Cheng, "A simple strategy to evaluate the frequency spectrum of a time series data with non-uniform intervals," *Transactions of the Aeronautical and Astronautical Society of the Republic of China*, vol. 36, no. 3, pp. 201–214, 2004.

[34] A. Capozzoli, C. Curcio, A. Liseno, and A. Riccardi, "Selecting parameters of type-3 NUFFTs to control accuracy in MoM methods," in *Antennas and*

Propagation (EuCAP), 2014 8th European Conference on, pp. 1157–1161, April 2014.

[35] J. Song, Q. H. Liu, S. Gewalt, G. Cofer, G. Johnson, and Q. H. Liu, "Least-square NUFFT methods applied to 2-D and 3-D radially encoded MR image reconstruction," *Biomedical Engineering, IEEE Transactions on*, vol. 56, pp. 1134–1142, April 2009.

[36] S. Aly, "Partially occluded pedestrian classification using histogram of oriented gradients and local weighted linear kernel support vector machine," *Computer Vision, IET*, vol. 8, no. 6, pp. 620–628, 2014.

[37] A. Youssef, T. Abdel-Galil, E. El-Saadany, and M. Salama, "Disturbance classification utilizing dynamic time warping classifier," *Power Delivery, IEEE Transactions on*, vol. 19, pp. 272–278, Jan 2004.

[38] R. Muscillo, S. Conforto, M. Schmid, P. Caselli, and T. D'Alessio, "Classification of motor activities through derivative dynamic time warping applied on accelerometer data," in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pp. 4930–4933, Aug 2007.

[39] Merriam-Webster.com, "Sample," 2016. `http://www.merriam-webster.com/dictionary/sample`.

[40] R. Keys, "Cubic convolution interpolation for digital image processing," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 29, pp. 1153–1160, Dec 1981.

[41] L. Wei and E. Keogh, "Semi-supervised time series classification," in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '06, (New York, NY, USA), pp. 748–753, ACM, 2006.

[42] L. Ye and E. Keogh, "Time series shapelets: A new primitive for data mining," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, (New York, NY, USA), pp. 947–956, ACM, 2009.

[43] X. Wang, A. Mueen, H. Ding, G. Trajcevski, P. Scheuermann, and E. Keogh, "Experimental comparison of representation methods and distance measures for time series data," *Data Mining and Knowledge Discovery*, vol. 26, no. 2, pp. 275–309, 2013.

[44] I. Stajduhar, B. Dalbelo-Basic, and N. Bogunovic, "Impact of censoring on learning Bayesian networks in survival modelling," *Artificial Intelligence in Medicine*, vol. 47, pp. 199–217, 2009.

[45] K. B. Howell, *Ordinary Differential Equations: An Introduction to the Fundamentals*. Hayden-McNeil Publishing, 2014–2015 edition, 2015.

[46] E. W. Kamen and B. S. Heck, *Fundamentals of Signals and Systems using the Web and Matlab*. Pearson Prentice Hall, 3rd ed., 2007.

[47] I. Bartolini, P. Ciaccia, and M. Patella, "WARP: Accurate retrieval of shapes using phase of Fourier descriptors and time warping distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, pp. 142–147, Jan. 2005.

[48] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-time Signal Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2nd ed., 1999.

[49] H. Ferhatosmanoglu, E. Tuncel, D. Agrawal, and A. El Abbadi, "Vector approximation based indexing for non-uniform high dimensional data sets," in *Proceedings of the Ninth International Conference on Information and Knowledge Management*, CIKM '00, (New York, NY, USA), pp. 202–209, ACM, 2000.

[50] M. M. M. Fuad, "Using differential evolution to set weights to segments with different information content in the piecewise aggregate approximation.," in *KES*, pp. 440–449, 2012.

[51] S. J. Russell and P. Norvig, *Artificial intelligence: A modern approach*. Prentice Hall series in artificial intelligence, Upper Saddle River: Prentice Hall, 3rd ed., 2010.

[52] Y. Chen, E. Keogh, B. Hu, N. Begum, A. Bagnall, A. Mueen, and G. Batista, "The ucr time series classification archive," July 2015. `www.cs.ucr.edu/~eamonn/time_series_data/`.

[53] K. A. Buza, *Fusion Methods for Time-Series Classification.* Peter Lang, 2011.

[54] M. M. M. Fuad, "Differential evolution versus genetic algorithms: Towards symbolic aggregate approximation of non-normalized time series," in *Proceedings of the 16th International Database Engineering & Applications Symposium*, IDEAS '12, (New York, NY, USA), pp. 205–210, ACM, 2012.

[55] M. Kokare, B. Chatterji, and P. Biswas, "Comparison of similarity metrics for texture image retrieval," in *TENCON 2003. Conference on Convergent Technologies for the Asia-Pacific Region*, vol. 2, pp. 571–575 Vol.2, Oct 2003.

[56] B. Mu and J. Yan, "An adaptive filtering technique for time series search," in *Business Intelligence and Financial Engineering, 2009. BIFE '09. International Conference on*, pp. 283–287, July 2009.

[57] Y. Cai and R. Ng, "Indexing spatio-temporal trajectories with chebyshev polynomials," in *Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data*, SIGMOD '04, (New York, NY, USA), pp. 599–610, ACM, 2004.

[58] M.-U.-S. Shameem and R. Ferdous, "An efficient k-means algorithm integrated with Jaccard distance measure for document clustering," in *Internet, 2009. AH-ICI 2009. First Asian Himalayas International Conference on*, pp. 1–6, Nov 2009.

[59] J. Wakerly, *Digital Design: Principles and Practices (4th Edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2005.

[60] G. Qian, S. Sural, Y. Gu, and S. Pramanik, "Similarity between Euclidean and cosine angle distance for nearest neighbor queries," in *Proceedings of the 2004 ACM symposium on Applied computing*, pp. 1232–1237, ACM, 2004.

[61] P. C. Mahalanobis, "On the generalized distance in statistics," *Proceedings of the National Institute of Sciences (Calcutta)*, vol. 2, pp. 49–55, 1936.

[62] J. Aach and G. M. Church, "Aligning gene expression time series with time warping algorithms," *Bioinformatics*, vol. 17, no. 6, pp. 495–508, 2001.

[63] I. Batal, H. Valizadegan, G. F. Cooper, and M. Hauskrecht, "A temporal pattern mining approach for classifying electronic health record data," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 4, no. 4, p. 63, 2013.

[64] I. Batal, H. Valizadegan, G. F. Cooper, and M. Hauskrecht, "A pattern mining approach for classifying multivariate temporal data," in *Bioinformatics and Biomedicine (BIBM), 2011 IEEE International Conference on*, pp. 358–365, IEEE, 2011.

[65] Y. Shahar, "A framework for knowledge-based temporal abstraction," *Artificial intelligence*, vol. 90, no. 1, pp. 79–133, 1997.

[66] M. Lichman, "UCI machine learning repository," 2013. http://archive.ics.uci.edu/ml.

[67] L. Trujillo, Y. Martínez, E. Galván-López, and P. Legrand, "Predicting problem difficulty for genetic programming applied to data classification," in *Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation*, GECCO '11, (New York, NY, USA), pp. 1355–1362, ACM, 2011.

[68] T. Hastie, *The elements of statistical learning: data mining, inference, and prediction*. Springer series in statistics, New York, NY: Springer, 2nd ed ed., 2009.

[69] E. Mayoraz and E. Alpaydın, "Support vector machine for multiclass classification," in *In Proceedings of International Workshop of Artificial Neural Networks (IWANN'99)*, 1999.

[70] A. Statnikov, C. F. Aliferis, I. Tsamardinos, D. Hardin, and S. Levy, "A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis," *Bioinformatics*, vol. 21, no. 5, pp. 631–643, 2005.

[71] O. Barrière, E. Lutton, and P.-H. Wuillemin, "Bayesian network structure learning using cooperative coevolution," in *Proceedings of the 11th Annual Conference on Genetic and Evolutionary Computation*, GECCO '09, (New York, NY, USA), pp. 755–762, ACM, 2009.

[72] N. Friedman, M. Linial, and I. Nachman, "Using Bayesian networks to analyze expression data," *Journal of Computational Biology*, vol. 7, pp. 601–620, 2000.

[73] R. E. Neapolitan, *Learning Bayesian networks*. Prentice Hall series in artificial intelligence, Upper Saddle River, NJ: Pearson Prentice Hall, 2004.

[74] "Bayesian networks and bayesian classifier software," *http://www.kdnuggets.com/software/bayesian.html*, 2016.

[75] "Genie and smile," *https://dslpitt.org/genie/*, 2013.

[76] M. J. Druzdzel, "SMILE: Structural modeling, inference, and learning engine and GeNIe: A development environment for graphical decision-theoretic models," in *Proceedings of the Sixteenth National Conference on Artificial Intelligence and the Eleventh Innovative Applications of Artificial Intelligence Conference Innovative Applications of Artificial Intelligence*, AAAI '99/IAAI '99, (Menlo Park, CA, USA), pp. 902–903, American Association for Artificial Intelligence, 1999.

[77] E. Ertin and L. C. Potter, "A method for sparse support vector regression," in *Defense and Security*, pp. 24–30, International Society for Optics and Photonics, 2005.

[78] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, pp. 273–297, Sept. 1995.

[79] V. Vapnik, *The nature of statistical learning theory*. Springer Science & Business Media, 2013.

[80] S. Yu, L.-C. Tranchevent, B. D. Moor, and Y. Moreau, *Kernel-based Data Fusion for Machine Learning: Methods and Applications in Bioinformatics and Text Mining*. Springer, 2011.

[81] C.-W. Hsu, C.-C. Chang, C.-J. Lin, *et al.*, "A practical guide to support vector classification," 2003.

[82] V. N. Vapnik, *Statistical learning theory*, vol. 1. Wiley New York, 1998.

[83] N. Cristianini and J. Shawe-Taylor, *Introduction to Support Vector Machines and other kernel-based learning methods*. Cambridge University Press, 2003.

[84] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of machine learning*. Cambridge, MA: MIT Press, 2012.

[85] Q. Wu, "The forecasting model based on wavelet ν-support vector machine," *Expert Syst. Appl.*, vol. 36, pp. 7604–7610, May 2009.

[86] B.-Y. Sun, Z.-H. Zhu, J. Li, and B. Linghu, "Combined feature selection and cancer prognosis using support vector machine regression," *IEEE/ACM Trans. Comput. Biol. Bioinformatics*, vol. 8, pp. 1671–1677, Nov. 2011.

[87] T. Farooq, A. Guergachi, and S. Krishnan, "Chaotic time series prediction using knowledge based Green's kernel and least-squares support vector machines," in *Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on*, pp. 373–378, Oct 2007.

[88] W. Qu, Y. He, and W. Qu, "Research on forecasting approach for complex time series based on support vector machines," in *Information Engineering and Computer Science (ICIECS), 2010 2nd International Conference on*, pp. 1–4, Dec 2010.

[89] S. Hartmann, "A competitive genetic algorithm for resource-constrained project scheduling," *Naval Research Logistics (NRL)*, vol. 45, no. 7, pp. 733–750, 1998.

[90] R. Herbrich, *Learning Kernel Classifiers: Theory and Algorithms*. Cambridge, MA, USA: MIT Press, 2001.

[91] E. Alpaydin, *Introduction to machine learning*. Cambridge, Mass: MIT Press, 2nd ed., 2010.

[92] S. Billings and W. Voon, "A prediction-error and stepwise-regression estimation algorithm for non-linear systems," *International Journal of Control*, vol. 44, no. 3, pp. 803–822, 1986.

[93] J. O. Rawlings, S. G. Pantula, and D. A. Dickey, *Applied regression analysis: a research tool*. Springer Science & Business Media, 2001.

[94] Mathworks, "stepwiselm," *https://www.mathworks.com/help/stats/stepwiselm.html*, 2016.

[95] C. E. Rasmussen, *Gaussian processes for machine learning*. Adaptive computation and machine learning, Cambridge, Mass: MIT Press, 2006.

[96] J. Quiñonero Candela and C. E. Rasmussen, "A unifying view of sparse approximate gaussian process regression," *J. Mach. Learn. Res.*, vol. 6, pp. 1939–1959, Dec. 2005.

[97] P. Baldi, *Bioinformatics: The Machine Learning Approach*. Adaptive computation and machine learning, Cambridge, Mass: MIT Press, 1998.

[98] A. P. Engelbrecht, *Computational intelligence: an introduction*. John Wiley & Sons, 2007.

[99] Mathworks, "Discrete convolution and the discrete Fourier transform," 2016. `http://www.mathworks.com/help/gads/computing-objective-functions.html`.

[100] M. Mitchell, *An introduction to genetic algorithms*. MIT press, 1998.

[101] J. H. Holland, *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. U Michigan Press, 1975.

[102] J. H. Holland, *Signals and boundaries: Building blocks for complex adaptive systems*. Mit Press, 2012.

[103] D. C. McKinney and M.-D. Lin, "Genetic algorithm solution of groundwater management models," *Water Resources Research*, vol. 30, no. 6, pp. 1897–1906, 1994.

[104] S. Nagendra, D. Jestin, Z. Gürdal, R. T. Haftka, and L. T. Watson, "Improved genetic algorithm for the design of stiffened composite panels," *Computers & Structures*, vol. 58, no. 3, pp. 543–555, 1996.

[105] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Computational and Structural Biotechnology Journal*, 2014.

[106] A. Mathew, M. Pandey, and N. Murthy, "Survival analysis: caveats and pitfalls," *European Journal of Surgical Oncology (EJSO)*, vol. 25, pp. 321–329, June 1999.

[107] O. Gevaert, F. D. Smet, D. Timmerman, Y. Moreau, and B. D. Moor, "Predicting the prognosis of breast cancer by integrating clinical and microarray data with bayesian networks," *Bioinformatics*, vol. 22, no. 14, pp. e184–e190, 2006.

[108] L. Chan, T. Chan, L. Cheng, and W. Mak, "Machine learning of patient similarity: A case study on predicting survival in cancer patient after locoregional

chemotherapy," in *Bioinformatics and Biomedicine Workshops (BIBMW), 2010 IEEE International Conference on*, pp. 467–470, Dec 2010.

[109] K. Park, A. Ali, D. Kim, Y. An, M. Kim, and H. Shin, "Robust predictive model for evaluating breast cancer survivability," *Engineering Applications of Artificial Intelligence*, vol. 26, no. 9, pp. 2194 – 2205, 2013.

[110] P. J. Lisboa and A. F. Taktak, "The use of artificial neural networks in decision support in cancer: A systematic review," *Neural Networks*, vol. 19, pp. 408–415, May 2006.

[111] J. M. Jerez-Aragonés, J. A. Gómez-Ruiz, G. Ramos-Jiménez, J. Muñoz Pérez, and E. Alba-Conejo, "A combined neural network and decision trees model for prognosis of breast cancer relapse," *Artif. Intell. Med.*, vol. 27, pp. 45–63, Jan. 2003.

[112] A.-L. Boulesteix and W. Sauerbrei, "Added predictive value of high-throughput molecular data to clinical data and its validation," *Briefings in Bioinformatics*, vol. 12, no. 3, pp. 215–229, 2011.

[113] L. Ohno-Machado, "Modeling medical prognosis: Survival analysis techniques," *Journal of Biomedical Informatics*, vol. 34, pp. 428–439, December 2001.

[114] P. Royston, "The lognormal distribution as a model for survival time in cancer, with an emphasis on prognostic factors," *Statistica Neerlandica*, vol. 55, no. 1, pp. 89–104, 2001.

[115] M.-H. Chen, J. G. Ibrahim, and D. Sinha, "A new joint model for longitudinal and survival data with a cure fraction," *Journal of Multivariate Analysis*, vol. 91, no. 1, pp. 18 – 34, 2004. Special Issue on Semiparametric and Nonparametric Mixed Models.

[116] I. Kononenko, "Machine learning for medical diagnosis: history, state of the art and perspective," *Artificial Intelligence in Medicine*, vol. 23, no. 1, pp. 89 – 109, 2001.

[117] B. Gan, C.-H. Zheng, and H. qiang Wang, "A survey of pattern classification-based methods for predicting survival time of lung cancer patients," in *Bioinformatics and Biomedicine (BIBM), 2014 IEEE International Conference on*, pp. 5–12, Nov 2014.

[118] A. Vellido and P. J. Lisboa, "Neural networks and other machine learning methods in cancer research," in *Computational and Ambient Intelligence*, pp. 964–971, Springer, 2007.

[119] S. Gupta, T. Tran, W. Luo, D. Phung, R. L. Kennedy, A. Broad, D. Campbell, D. Kipp, M. Singh, M. Khasraw, L. Matheson, D. M. Ashley, and S. Venkatesh, "Machine-learning prediction of cancer survival: A retrospective study using electronic administrative records and a cancer registry," *BMJ Open*, vol. 4, no. 3, 2014.

[120] T. Hothorn, P. Bühlmann, S. Dudoit, A. Molinaro, and M. J. Van Der Laan, "Survival ensembles," *Biostatistics*, vol. 7, no. 3, pp. 355–373, 2006.

[121] J. Kalderstam, P. Eden, P.-O. Bendahl, C. Strand, and M. Ferno, "Training artificial neural networks directly on the concordance index for censored data using genetic algorithms," *Artificial Intelligence in Medicine*, vol. 58, pp. 125–132, 2013.

[122] D. Delen, G. Walker, and A. Kadam, "Predicting breast cancer survivability: a comparison of three data mining methods," *Artificial Intelligence in Medicine*, vol. 34, pp. 113–127, 2005.

[123] C. A. Stone, E. Tiernan, and B. A. Dooley, "Prospective validation of the palliative prognostic index in patients with cancer," *Journal of Pain and Symptom Management*, vol. 35, no. 6, p. 617622, 2008.

[124] C. W. Hug and P. Szolovits, "Icu acuity: real-time models versus daily models.," in *AMIA*, 2009.

[125] P. Royston, "The lognormal distribution as a model for survival time in cancer, with an emphasis on prognostic factors," *Statistica Neerlandica*, vol. 55, no. 1, pp. 89–104, 2001.

[126] I. Batal, L. Sacchi, R. Bellazzi, and M. Hauskrecht, "A temporal abstraction framework for classifying clinical temporal data," in *AMIA Annual Symposium Proceedings*, vol. 2009, p. 29, American Medical Informatics Association, 2009.

[127] W. Xiao-lu, L. Jian, and L. Jian-jun, "Wavelet transform and pso support vector machine based approach for time series forecasting," in *Artificial Intelligence and Computational Intelligence, 2009. AICI '09. International Conference on*, vol. 1, pp. 46–50, Nov 2009.

[128] N. Sapankevych and R. Sankar, "Constrained motion particle swarm optimization and support vector regression for non-linear time series regression and prediction applications," in *Machine Learning and Applications (ICMLA), 2013 12th International Conference on*, vol. 2, pp. 473–477, Dec 2013.

[129] H. Lei and B. Sun, "A study on the dynamic time warping in kernel machines," in *Signal-Image Technologies and Internet-Based System, 2007. SITIS '07. Third International IEEE Conference on*, pp. 839–845, Dec 2007.

[130] X. Yang, G. Zhang, J. Lu, and J. Ma, "A kernel fuzzy c-means clustering-based fuzzy support vector machine algorithm for classification problems with outliers or noises," *Fuzzy Systems, IEEE Transactions on*, vol. 19, pp. 105–115, Feb 2011.

[131] H. Chao, H. Li-li, and H. Ting-ting, "Financial time series forecasting based on wavelet kernel support vector machine," in *Natural Computation (ICNC), 2012 Eighth International Conference on*, pp. 79–83, May 2012.

[132] S. Lessmann, R. Stahlbock, and S. Crone, "Genetic algorithms for support

vector machine model selection," in *Neural Networks, 2006. IJCNN '06. International Joint Conference on*, pp. 3063–3069, 2006.

[133] I. Batal, D. Fradkin, J. Harrison, F. Moerchen, and M. Hauskrecht, "Mining recent temporal patterns for event detection in multivariate time series data," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 280–288, ACM, 2012.

[134] J. Lin, E. gh, L. Wei, and S. Lonardi, "Experiencing sax: A novel symbolic representation of time series," *Data Min. Knowl. Discov.*, vol. 15, pp. 107–144, Oct. 2007.

[135] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra, "Locally adaptive dimensionality reduction for indexing large time series databases," *ACM SIGMOD Record*, vol. 30, no. 2, pp. 151–162, 2001.

[136] E. Keogh, S. Chu, D. Hart, and M. Pazzani, "Segmenting time series: A survey and novel approach," *Data mining in time series databases*, vol. 57, pp. 1–22, 2004.

[137] E. Fuchs, T. Gruber, J. Nitschke, and B. Sick, "Online segmentation of time series based on polynomial least-squares approximations," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 12, pp. 2232–2245, 2010.

[138] H. Guo, X. Liu, and L. Song, "Dynamic programming approach for segmentation of multivariate time series," *Stochastic environmental research and risk assessment*, vol. 29, no. 1, pp. 265–273, 2015.

[139] L. Karamitopoulos and G. Evangelidis, "Current trends in time series representation," in *Proc. 11th Panhellenic Conference on Informatics*, pp. 217–226, 2007.

[140] A. Gionis and H. Mannila, "Finding recurrent sources in sequences," in *Proceedings of the Seventh Annual International Conference on Research in Computational Molecular Biology*, RECOMB '03, (New York, NY, USA), pp. 123–130, ACM, 2003.

[141] F.-L. Chung, T.-C. Fu, V. Ng, and R. W. Luk, "An evolutionary approach to pattern-based time series segmentation," *Evolutionary Computation, IEEE Transactions on*, vol. 8, no. 5, pp. 471–489, 2004.

[142] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series.," in *KDD workshop*, vol. 10, pp. 359–370, Seattle, WA, 1994.

[143] M. D. Sacchi, T. J. Ulrych, and C. J. Walker, "Interpolation and extrapolation using a high-resolution discrete fourier transform," *Signal Processing, IEEE Transactions on*, vol. 46, no. 1, pp. 31–38, 1998.

[144] J. Fessler and B. Sutton, "Nonuniform fast Fourier transforms using min-max

interpolation," *Signal Processing, IEEE Transactions on*, vol. 51, pp. 560–574, Feb 2003.

[145] D. L. Russell, "Discrete convolution and the discrete Fourier transform," 2014. `www.math.vt.edu/people/russell/m2k_opm_dfour3.pdf`.

[146] H. Shimodaira, K. Noma, M. Nakai, and S. Sagayama, "Dynamic time-alignment kernel in support vector machine," *Advances in neural information processing systems*, vol. 14, p. 921, 2002.

Appendix A

Background

This appendix contains extends on the theory of sampling (Appendix A.1) and Fourier (Appendix A.1.1). Appendix A.2 explains the implementation of dynamic time warping.

A.1 Sampling Theory

With the spacing between samples described, the Whittaker-Kotel'nikov-Shannon (WSK) sampling theory is formed. Using WSK with the time series $\mathbf{S}$, the signal $\hat{S}(t)$

can be determined. The reconstructed signal $\hat{S}(t)$ is described by the summation [21],

$$\hat{S}(t) = \sum_{n=-\infty}^{\infty} \mathbf{S}\left(\frac{n\pi}{\Omega}\right) \frac{\sin(\Omega t - n\pi)}{\Omega t - n\pi} = \sin(\Omega t) \sum_{n=-\infty}^{\infty} \mathbf{S}\left(\frac{n\pi}{\Omega}\right) \frac{(-1)^n}{\Omega t - n\pi}. \tag{A.1}$$

In the WSK Theorem, the ability to acquire uniformly spaced observations from $\mathbf{S}$ is needed. When there is no value obtainable at the desired time point additional considerations must be made. Then to obtain uniformly spaced observations one approach is interpolation. Using surrounding observations interpolation estimates the missing information needed to estimate an observation at that point of time. There are several interpolation approaches, often requiring a function to describe the signal to estimate the observation at the desired time point.

An alternative to interpolation is gridding. The gridding kernel properties share the same principles as those used in Support Vector Machines (see section 2.3.3.1 for more background on kernels). For a kernel, K, will act as one of the input functions when performing the convolution. The convolution is defined as,

$$(K * \mathbf{S})(t) = \int_0^t K(\tau)\mathbf{S}(t-\tau)d\tau. \tag{A.2}$$

Within the convolution, the integral contains the product operation of $K(\tau)\mathbf{S}(y)$, which is constrained to $\tau + y = t$ [145]. Similarly in the discrete form, the kernel

takes the role of one of the vectors with the sample considered as the other vector in the general discrete convolution form,

$$(K * \mathbf{S})_k = \sum_{j=0}^{N-1} K_{k-j} s_j. \tag{A.3}$$

With the convolution's ability to resolve the non-continuous behavior of a sample, taking the integral or summation is a task which may still be very complex or difficult. One of the properties of convolution which simplifies the computation is the utilization of a transform. The Laplace Transform of a convolution results in the product of the Laplace Transform of each function individually [45]. This convolution property also is applicable to the FT. This convolution property then allows for the convolution of a kernel, K, with a sample, $\mathbf{S}$, to be defined as,

$$\mathcal{F}[(K * \mathbf{S})(t)] = \mathcal{F}(K)(\Omega)\mathcal{F}(\mathbf{S})(\Omega). \tag{A.4}$$

A.1.1 Fourier Representation

The function form of the FT for the function g(t) is,

$$G(\Omega) = \int_{-\infty}^{\infty} g(t)\, e^{-i\Omega t} dt. \tag{A.5}$$

For the discrete form, the integral is replaced with a summation. The FT is then known as the Discrete Time Fourier Transform (DTFT) [46].

For a sample, **S**, the DTFT is determined,

$$S(\Omega) = \sum_{n=0}^{N-1} s[n] e^{-i\Omega n}, \tag{A.6}$$

and the bound of the summation is tightened to the number of samples that exists for the sample [46] compared to the infinite endpoints of the integral form.

In the notation for both transforms, the periodic function $e^{-i\Omega n}$ is used, with i being the representation of $\sqrt{-1}$. The DTFT is made into the DFT with the evaluation being made for the frequency, using $\Omega = \frac{2\pi k}{N}$ [46]. This substitution results in the DFT with N samples being [46],

$$S_k = \sum_{n=0}^{N-1} s[n] e^{-\frac{i2\pi kn}{N}}. \tag{A.7}$$

Once the function is periodic, the return of the sample to the time domain is needed to allow for the results to be used, this is done using the Inverse Fourier Transform (IFT) or Inverse Discrete Fourier Transform (IDFT).

The IFT is performed with applying an integral,

$$S(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\Omega) e^{i\Omega t} d\Omega. \tag{A.8}$$

The IDFT is done with an integral (presented in DFTF form), which is possible since the DFT transformed the series into a continuous function, [46],

$$s[n] = \frac{1}{2\pi} \int_{0}^{2\pi} S(\Omega) e^{i\Omega n} d\Omega. \tag{A.9}$$

A.2 Dynamic Time Warping

To implement dynamic time warping (DTW) [62], the comparison is not direct comparison between observations as done with Euclidean Distance, but the corresponding observations in the sample is being compared to all the observations in the other sample. In Fig. A.1, two time series **S** and **R** are compared by both DTW and Euclidean Distance metrics with the green lines representing the selected distanced used in the similarity calculation with the lines extending between the series representing the distances calculated.

The process for DTW is recursive, comparing two time series at a time. Using the time series shown in Fig. A.1, sample **S** is being compared with **R**. In both series

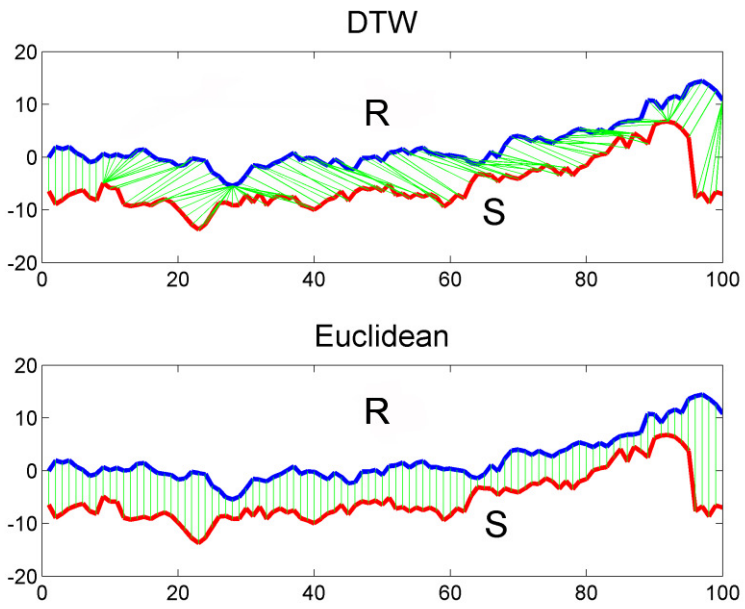


Figure A.1: Visual comparison of distance calculations comparing time series **R** and series **S** with DTW and Euclidean distance metrics [2]. © 2012 P. Montalto, M. Aliotta, A. Cannata, C. Cassisi, and A. Pulvirenti. Adapted from "Advances in Data Mining Knowledge Discovery and Applications", under CC by 3.0 license. Available from: doi.org/10.5772/49941.

there are m observations, allowing for i and j to index each series from $\{0, 1, ...m-1\}$.

DTW has been explored for similarity detection, recognition of activity [38], and power disturbances [37]. There is caution with the application of DTW, the triangle inequality is not satisfied, thus creating needs for lower bounding methodologies when indexing to prevent complications [56].

Forecasting has utilized dynamic kernels based on DTW with SVM [6]. DTW has been merged with a piecewise approach, with equal sized frames on American Sign Language data set from the University of California Irvine (UCI) and cylinder bell

and funnel set [17]. Other considerations of DTW incorporating in to kernels have also been done [129, 146].

Appendix B

Support Vector Machines

Classification Results

The tables contain the full results from the Support Vector Machines (SVM) aspect of Chapter 3. With Table B.1 presenting for the SVM with the linear kernel and t_1^*, and Table B.2 with t_2^*. The radial basis function (RBF) kernel using t_1^* is presented in Table B.3, and t_2^* in Table B.4.

Table B.1
Linear kernel SVM with t_1^*.

# of Diffs	Splines	Samples	% Accuracy	
			7 Day	14 Day
0	no	0	34.09±3.1	34.09±3.1
0	yes	0	37.37±2.5	37.37±2.5
1	no	0	34.40±2.8	34.40±2.8
1	yes	0	37.60±2.7	37.60±2.7
2	yes	0	39.13±3.3	39.13±3.3
2	no	0	36.00±2.6	36.54±2.4
0	no	1	34.25±3.2	34.25±3.2
0	yes	1	37.68±2.7	37.68±2.7
1	no	1	35.47±3.2	35.47±3.0
1	yes	1	39.13±4.3	39.13±4.3
2	no	1	35.77±2.6	35.77±2.6
2	yes	1	38.68±3.1	38.68±3.0
0	no	2	35.70±2.1	36.08±2.4
0	yes	2	38.06±2.4	38.67±2.8
1	no	2	37.22±2.8	37.76±2.3
1	yes	2	38.52±4.2	39.59±3.7
2	no	2	36.54±2.4	36.54±2.4
2	yes	2	37.99±3.6	39.29±4.0
0	no	3	36.00±2.3	36.31±2.3
0	yes	3	38.22±2.6	38.98±3.1
1	no	3	37.76±2.6	37.83±2.9
1	yes	3	37.30±4.1	39.66±3.5
2	no	3	37.15±2.7	36.92±2.4
2	yes	3	39.06±4.6	38.37±4.3

Table B.2
Linear kernel SVM with t_2^*.

# of Diffs	Splines	Samples	% Accuracy	
			7 Day	**14 Day**
0	no	0	34.09±3.2	34.09±3.2
0	yes	0	36.30±3.2	36.30±3.2
1	no	0	33.94±3.1	33.94±3.1
1	yes	0	36.38±4.2	36.38±4.2
2	yes	0	38.13±3.2	38.13±3.2
2	no	0	35.46±3.3	35.46±3.3
0	no	1	34.86±2.1	34.86±2.1
0	yes	1	38.13±3.2	35.85±2.8
1	no	1	35.77±3.1	35.77±3.1
1	yes	1	38.06±3.6	38.06±3.6
2	no	1	36.23±2.0	36.23±2.0
2	yes	1	38.36±3.8	38.36±3.8
0	no	2	35.69±3.3	36.68±3.3
0	yes	2	36.07±3.5	36.53±4.1
1	no	2	37.37±4.1	36.91±3.8
1	yes	2	37.30±3.2	37.75±4.0
2	no	2	36.83±2.7	37.45±3.3
2	yes	2	38.36±3.4	38.82±3.9
0	no	3	36.68±3.7	36.83±4.2
0	yes	3	35.84±3.9	36.84±2.8
1	no	3	37.30±4.1	36.84±3.8
1	yes	3	38.06±3.4	37.52±3.9
2	no	3	37.37±3.1	36.84±3.7
2	yes	3	38.74±3.9	38.29±3.6

Table B.3
RBF kernel SVM with t_1^*.

# of Diffs	Splines	Samples	% Accuracy	
			7 Day	**14 Day**
0	no	0	34.55±3.0	34.55±3.0
0	yes	0	36.23±3.7	36.23±3.7
1	no	0	34.10±3.7	34.10±3.7
1	yes	0	33.56±2.5	33.56±2.5
2	yes	0	34.25±3.1	34.25±3.1
2	no	0	33.87±4.6	33.87±4.6
0	no	1	34.55±4.3	34.55±4.2
0	yes	1	34.48±4.1	34.48±4.1
1	no	1	33.33±4.2	34.02±4.1
1	yes	1	32.72±3.3	32.72±3.3
2	no	1	34.10±3.5	34.10±3.5
2	yes	1	33.64±3.5	33.18±3.3
0	no	2	35.77±3.7	36.53±4.1
0	yes	2	34.17±3.7	34.55±3.7
1	no	2	33.33±4.2	33.25±4.4
1	yes	2	32.95±3.3	32.87±3.3
2	no	2	34.02±3.8	34.10±3.8
2	yes	2	33.64±3.5	33.64±3.5
0	no	3	35.47±4.4	36.23±3.9
0	yes	3	34.40±3.7	34.55±3.9
1	no	3	34.17±3.8	33.02±4.1
1	yes	3	32.57±2.8	32.72±3.3
2	no	3	34.17±3.8	34.10±4.2
2	yes	3	33.64±3.5	33.56±3.4

Table B.4
RBF kernel SVM with t_2^*.

# of Diffs	Splines	Samples	% Accuracy	
			7 Day	14 Day
0	no	0	34.09±2.7	34.17±3.5
0	yes	0	35.23±3.2	35.84±3.0
1	no	0	35.31±3.1	35.31±3.1
1	yes	0	34.32±3.9	34.32±3.9
2	yes	0	34.32±2.5	34.02±2.6
2	no	0	35.01±3.3	34.86±2.1
0	no	1	33.26±2.2	33.94±2.5
0	yes	1	34.32±1.9	34.17±2.2
1	no	1	34.71±3.0	35.62±3.8
1	yes	1	33.03±1.8	33.26±3.8
2	no	1	34.71±3.0	34.48±2.1
2	yes	1	33.03±1.8	33.18±2.0
0	no	2	33.79±2.3	33.33±1.6
0	yes	2	35.08±1.7	34.71±2.4
1	no	2	35.47±3.1	35.39±3.4
1	yes	2	33.57±4.0	33.56±4.5
2	no	2	34.94±3.2	33.26±2.0
2	yes	2	33.48±2.2	33.56±2.2
0	no	3	34.10±3.4	34.48±2.4
0	yes	3	34.63±2.4	34.55±2.6
1	no	3	35.01±2.8	35.09±3.2
1	yes	3	33.26±4.2	33.41±4.5
2	no	3	35.01±2.8	34.86±2.9
2	yes	3	33.48±2.1	33.48±2.1

Appendix C

Regression Length of Survival

This appendix contains extended results on the length of survival prognosis using regression to complement the work presented in Chapter 4. The summary of the best performing models with the complete data set is presented in Table C.1 and the tables that follow are full model explorations.

C.1 Reduced Data Set Summary

Table C.1
BEST PERFORMING REGRESSION MODELS COMPLETE SET†.

Evaluation Method	**Num Samples**	**Features Included**	**Median Absolute Error (Days)**	**Median Relative Error**
Linear	1, 7 day	No Diffs, Splines	82.14	0.8469
Quadratic	0	No Diffs, Splines	86.65	0.9070
GP	1, 14 day	No Diffs, Splines	83.48	0.8910
SVR- Linear	4, 7 day	No Diffs, No Splines	50.41	0.7132
SVR- RBF	2, 7 day	No Diffs, No Splines	60.86	0.7701
Linear	5, 7 day	2 Diffs, Splines	84.38	0.8606
Quadratic	0	1 Diffs, Splines	91.29	0.9620
GP	0	No Diffs, Splines	82.63	0.9478
SVR- Linear	1, 14 day	2 Diffs, No Splines	47.41	0.7146
SVR- RBF	0	1 Diffs, Splines	43.38	0.7778

†Above the triple line is t_1^* and blow is t_2^*.

C.2 t_1^*

The following tables present the results of the regression evaluations for the complete DS1 data set using t_1^*. The order of the tables is linear regression, quadratic regression, Gaussian Process regression, SVR with a linear kernel and finally the SVR with the radial basis kernel for the absolute error. This order is then repeated for the relative error.

Table C.2
Linear regression with t_1^* absolute error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Abs. Error**	**Mean Abs. Error**	**Med. Abs. Error**	**Mean Abs. Error**
0	0	1	89.3	118.2±108.9	89.8	117.7±108.9
0	0	2	89.7	118.4±108.8	90.1	118.3±108.7
0	0	3	90.2	118.7±108.7	89.5	118.4±109.1
0	0	4	89.4	118.9±109.0	90.3	118.6±108.9
0	0	5	90.3	118.9±108.9	90.8	118.7±108.8
0	1	0	82.2	114.6±107.9	82.2	114.6±107.9
0	1	1	82.1	114.2±108.1	82.4	114.1±108.1
0	1	2	83.8	114.3±108.1	83.1	114.4±108.0
0	1	3	83.2	114.6±108.0	83.1	114.5±108.1
0	1	4	83.7	114.9±108.0	83.0	114.7±108.0
0	1	5	83.3	114.9±108.0	82.9	114.7±108.0
1	0	0	89.0	117.4±108.5	89.0	117.4±108.5
1	0	1	87.4	116.9±108.4	87.7	116.5±108.4
1	0	2	88.4	117.0±108.5	88.5	116.9±108.4
1	0	3	89.9	117.4±108.5	86.9	117.1±108.7
1	0	4	90.2	117.8±108.6	87.7	117.2±108.5
1	0	5	89.3	117.7±108.6	87.6	117.3±108.4
1	1	0	82.3	114.3±107.7	82.3	114.3±107.7
1	1	1	82.4	114.1±107.8	82.6	114.0±107.8
1	1	2	82.7	114.2±107.8	83.5	114.2±107.8
1	1	3	83.2	114.4±107.9	82.9	114.3±108.0
1	1	4	83.5	114.7±107.8	83.1	114.3±107.9
1	1	5	84.5	114.7±107.8	83.3	114.4±107.9
2	0	0	85.6	116.6±108.8	85.6	116.6±108.8
2	0	1	86.0	116.2±108.7	86.4	116.0±108.7
2	0	2	87.3	116.3±108.8	87.0	116.3±108.7
2	0	3	87.0	116.5±108.9	86.5	116.0±109.1
2	0	4	87.1	116.6±108.9	87.0	115.9±109.0
2	0	5	87.2	116.5±108.9	85.5	116.2±108.8
2	1	0	83.5	114.2±108.2	83.5	114.2±108.2
2	1	1	82.6	114.1±108.3	82.0	114.2±108.2
2	1	2	82.9	114.2±108.4	83.2	114.3±108.3
2	1	3	83.5	114.3±108.4	83.4	114.2±108.4
2	1	4	83.6	114.5±108.3	84.2	114.3±108.3
2	1	5	83.3	114.5±108.3	82.5	114.5±108.3

Table C.3
Quadratic regression with t_1^* absolute error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Abs. Error**	**Mean Abs. Error**	**Med. Abs. Error**	**Mean Abs. Error**
0	0	1	91.2	117.9±106.5	92.3	117.9±107.1
0	0	2	97.3	119.4±106.3	95.3	119.5±107.1
0	0	3	96.4	120.5±106.6	96.2	120.4±107.0
0	0	4	100.0	122.9±107.8	98.5	122.0±106.2
0	0	5	99.6	125.6±107.4	101.2	125.1±106.8
0	1	0	86.7	125.2±167.4	86.7	125.2±167.4
0	1	1	88.8	127.2±171.1	90.8	128.3±182.2
0	1	2	92.3	126.7±133.6	92.7	133.6±196.5
0	1	3	97.1	130.3±136.9	97.9	130.3±128.4
0	1	4	100.3	142.3±201.6	100.8	141.6±222.9
0	1	5	107.9	162.8±348.8	105.6	150.7±267.4
1	0	0	87.7	117.9±109.8	87.7	117.9±109.8
1	0	1	90.3	118.3±107.7	90.9	118.5±108.3
1	0	2	91.8	119.6±107.9	93.4	121.7±108.8
1	0	3	95.5	122.6±109.5	96.7	124.0±109.4
1	0	4	98.7	126.4±110.9	98.9	126.7±110.0
1	0	5	100.9	132.3±118.2	102.8	130.9±112.4
1	1	0	85.8	130.1±192.9	85.8	130.1±192.9
1	1	1	91.5	128.8±136.7	94.7	131.5±147.2
1	1	2	95.5	141.7±232.4	93.7	141.0±215.4
1	1	3	97.5	142.2±189.2	99.5	151.3±278.0
1	1	4	108.5	171.6±391.8	107.1	151.7±189.1
1	1	5	119.5	181.0±323.3	116.7	201.7±580.3
2	0	0	84.3	116.7±110.2	84.3	116.7±110.2
2	0	1	87.3	118.4±109.1	87.7	118.3±109.3
2	0	2	92.3	121.9±108.7	91.0	121.0±109.4
2	0	3	95.3	124.9±110.6	94.0	124.9±111.0
2	0	4	102.2	132.3±117.9	99.2	130.2±113.7
2	0	5	107.1	140.9±126.3	103.7	135.8±120.6
2	1	0	84.5	135.6±246.5	84.5	135.6±246.5
2	1	1	92.4	140.1±226.9	92.5	140.4±236.9
2	1	2	97.5	152.2±292.8	99.9	138.5±157.9
2	1	3	101.9	153.8±224.8	100.4	152.1±207.6
2	1	4	113.8	171.3±291.4	112.9	169.5±308.3
2	1	5	122.7	190.5±300.3	125.1	181.5±260.0

Table C.4
Gaussian Process regression with t_1^* absolute error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Abs. Error**	**Mean Abs. Error**	**Med. Abs. Error**	**Mean Abs. Error**
0	0	1	83.5	116.2±106.5	84.3	116.8±107.3
0	0	2	82.6	115.9±106.4	84.8	117.0±107.1
0	0	3	83.8	116.4±106.5	84.9	117.0±106.9
0	0	4	84.5	116.5±106.5	84.6	116.9±107.0
0	0	5	84.6	116.5±106.5	84.8	116.8±107.2
0	1	0	83.9	115.0±107.7	83.9	115.0±107.7
0	1	1	84.0	114.9±108.0	83.5	114.5±108.0
0	1	2	82.1	115.1±107.5	83.9	115.1±107.8
0	1	3	82.5	115.3±107.3	83.5	115.5±107.4
0	1	4	82.6	115.5±107.3	82.8	115.4±107.3
0	1	5	82.4	115.5±107.2	82.5	115.6±107.3
1	0	0	83.4	115.9±108.4	83.6	115.5±108.9
1	0	1	84.4	116.1±108.2	83.9	114.7±108.5
1	0	2	83.4	116.0±108.1	83.8	114.9±108.3
1	0	3	84.6	116.4±107.8	83.4	115.3±108.2
1	0	4	84.5	116.5±107.8	83.5	115.2±108.1
1	0	5	84.6	116.5±107.7	83.3	115.3±108.9
1	1	0	83.5	114.5±107.5	83.5	114.5±107.5
1	1	1	81.6	114.3±107.7	81.8	114.2±107.7
1	1	2	83.3	114.5±107.7	82.9	114.6±107.6
1	1	3	82.8	114.7±107.6	83.4	114.9±107.6
1	1	4	82.5	114.9±107.6	83.4	114.9±107.4
1	1	5	82.6	115.0±107.4	83.3	115.0±107.4
2	0	0	83.6	115.5±108.9	83.4	115.9±108.4
2	0	1	83.2	114.7±108.4	84.9	116.1±108.5
2	0	2	83.9	114.9±108.2	85.3	116.3±108.4
2	0	3	84.0	115.1±108.1	85.5	116.7±108.2
2	0	4	83.9	115.3±108.1	85.0	116.6±108.1
2	0	5	83.8	115.4±108.0	84.4	116.4±108.1
2	1	0	79.1	114.0±108.0	79.1	114.0±108.0
2	1	1	80.6	113.8±108.0	79.7	113.9±108.0
2	1	2	81.2	114.0±107.9	79.7	114.1±107.9
2	1	3	80.9	114.2±107.8	80.9	114.3±107.8
2	1	4	80.6	114.3±107.8	80.9	114.3±107.7
2	1	5	81.0	114.4±107.8	81.6	114.4±107.7

Table C.5
SVR linear regression with t_1^* absolute error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Abs. Error**	**Mean Abs. Error**	**Med. Abs. Error**	**Mean Abs. Error**
0	0	1	50.2	103.0±135.4	48.1	104.4±141.4
0	0	2	51.2	103.2±135.4	48.0	104.5±141.0
0	0	3	50.6	103.1±135.5	48.0	104.5±141.2
0	0	4	50.4	103.6±136.0	47.6	104.5±141.8
0	0	5	50.1	103.5±136.2	47.8	104.6±142.1
0	1	0	48.0	103.6±136.6	48.0	103.6±136.6
0	1	1	48.9	103.2±136.4	49.1	103.2±136.6
0	1	2	48.8	103.3±136.4	48.3	103.4±136.6
0	1	3	49.0	103.3±136.3	49.3	103.4±136.1
0	1	4	48.3	103.6±137.1	48.8	103.4±137.0
0	1	5	47.6	103.5±137.6	48.2	103.4±137.6
1	0	0	48.5	104.3±139.8	61.6	108.5±129.1
1	0	1	48.1	103.8±139.0	47.3	103.7±139.3
1	0	2	47.9	103.8±139.5	49.3	103.8±138.8
1	0	3	49.4	104.0±138.9	47.4	103.9±139.1
1	0	4	48.9	104.2±139.5	47.8	103.8±139.3
1	0	5	48.1	104.3±140.1	47.1	103.9±139.7
1	1	0	48.9	103.5±135.9	48.8	103.5±135.6
1	1	1	50.2	103.0±135.8	49.5	103.0±135.5
1	1	2	51.1	103.3±135.5	49.8	103.3±135.6
1	1	3	50.7	103.1±135.4	49.1	103.3±135.3
1	1	4	50.7	103.6±135.8	48.9	103.2±135.8
1	1	5	50.2	103.5±136.2	48.7	103.1±136.1
2	0	0	49.5	104.1±137.9	49.5	104.1±137.9
2	0	1	50.5	103.8±137.4	49.5	103.8±137.4
2	0	2	50.1	103.7±137.5	49.3	103.9±137.7
2	0	3	49.5	103.8±137.1	49.5	103.8±137.2
2	0	4	50.1	104.1±137.5	49.2	103.7±137.6
2	0	5	50.3	103.9±138.0	48.5	103.7±137.7
2	1	0	50.6	103.4±135.2	50.6	103.4±135.2
2	1	1	50.8	103.2±135.1	50.5	103.4±135.1
2	1	2	51.5	103.3±135.2	50.5	103.3±135.0
2	1	3	50.6	103.3±135.0	49.9	103.1±135.2
2	1	4	51.6	103.3±135.3	50.7	103.3±135.4
2	1	5	51.0	103.4±135.7	50.0	103.2±135.9

Table C.6
SVR with RBF kernel regression with t_1^* absolute error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Abs. Error**	**Mean Abs. Error**	**Med. Abs. Error**	**Mean Abs. Error**
0	0	1	59.8	108.3±130.3	58.8	107.9±131.9
0	0	2	60.9	109.0±129.4	61.2	109.2±132.1
0	0	3	62.0	110.5±132.3	61.2	109.8±132.9
0	0	4	61.7	109.7±130.5	60.8	108.8±132.4
0	0	5	58.3	107.8±131.9	58.2	108.0±133.8
0	1	0	59.7	108.7±132.0	75.9	117.0±126.7
0	1	1	62.2	109.0±129.5	61.4	108.8±129.4
0	1	2	63.8	110.5±129.8	63.2	110.3±130.5
0	1	3	64.0	111.0±131.9	63.2	110.7±132.1
0	1	4	63.6	110.1±129.4	62.7	109.1±130.2
0	1	5	61.6	109.1±130.8	61.4	108.9±131.1
1	0	0	54.1	73.5±82.7	59.9	109.7±133.2
1	0	1	61.6	109.5±129.0	61.6	108.5±129.1
1	0	2	62.9	110.6±129.7	64.6	110.6±129.5
1	0	3	66.6	112.0±128.9	65.7	111.4±129.1
1	0	4	65.3	111.4±128.1	64.0	109.8±128.3
1	0	5	62.8	110.8±130.8	62.5	109.8±130.5
1	1	0	50.9	74.0±82.6	63.5	110.0±130.1
1	1	1	62.8	109.5±127.8	62.8	109.1±127.9
1	1	2	64.8	111.1±129.1	66.4	111.8±129.1
1	1	3	66.9	111.6±128.3	66.2	111.7±128.3
1	1	4	66.0	111.0±126.8	65.6	110.2±127.1
1	1	5	63.8	110.6±129.8	63.3	110.3±129.2
2	0	0	65.5	111.4±131.2	81.7	120.6±127.6
2	0	1	65.6	110.1±126.4	64.7	110.1±126.5
2	0	2	68.7	111.4±128.2	65.7	110.7±127.3
2	0	3	67.7	112.8±129.9	67.1	112.6±129.5
2	0	4	63.6	108.6±127.7	64.4	108.9±127.6
2	0	5	69.5	113.7±128.1	68.4	112.8±128.1
2	1	0	66.9	111.5±130.8	82.2	119.5±127.3
2	1	1	66.8	110.2±125.0	66.9	109.7±125.2
2	1	2	69.5	111.3±127.5	66.8	110.5±127.3
2	1	3	65.9	112.0±129.7	66.4	111.7±129.8
2	1	4	64.3	108.7±127.1	65.4	108.8±126.8
2	1	5	69.9	113.8±127.6	69.1	112.8±127.9

Table C.7
Linear regression with t_1^* relative error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Rel. Error**	**Mean Rel. Error**	**Med. Rel. Error**	**Mean Rel. Error**
0	0	1	1.01	4.61±12.6	0.99	4.57±12.4
0	0	2	1.00	4.61±12.6	1.01	4.58±12.5
0	0	3	1.01	4.62±12.7	1.01	4.61±12.7
0	0	4	1.01	4.62±12.6	1.01	4.64±12.8
0	0	5	1.01	4.61±12.6	1.02	4.66±13.0
0	1	0	0.88	4.42±11.7	0.88	4.42±11.7
0	1	1	0.85	4.40±11.8	0.85	4.37±11.7
0	1	2	0.87	4.40±11.8	0.86	4.38±11.7
0	1	3	0.86	4.41±11.9	0.86	4.41±11.8
0	1	4	0.88	4.41±11.8	0.89	4.42±11.8
0	1	5	0.87	4.40±11.8	0.87	4.44±12.0
1	0	0	0.97	4.46±11.8	0.97	4.46±11.8
1	0	1	0.95	4.43±11.6	0.97	4.39±11.4
1	0	2	0.97	4.43±11.7	0.98	4.41±11.6
1	0	3	0.99	4.45±11.8	0.98	4.44±11.9
1	0	4	1.01	4.45±11.8	0.95	4.45±11.9
1	0	5	0.99	4.43±11.7	0.97	4.46±12.0
1	1	0	0.88	4.34±11.4	0.88	4.34±11.4
1	1	1	0.86	4.32±11.4	0.86	4.30±11.3
1	1	2	0.88	4.32±11.4	0.89	4.31±11.4
1	1	3	0.87	4.33±11.5	0.86	4.33±11.5
1	1	4	0.87	4.34±11.5	0.86	4.34±11.5
1	1	5	0.88	4.32±11.5	0.86	4.36±11.6
2	0	0	0.97	4.39±11.1	0.97	4.39±11.1
2	0	1	0.97	4.35±11.0	0.97	4.34±11.0
2	0	2	1.00	4.35±11.1	0.95	4.37±11.2
2	0	3	0.99	4.36±11.2	0.91	4.39±11.4
2	0	4	0.99	4.37±11.1	0.92	4.40±11.6
2	0	5	0.95	4.34±11.1	0.94	4.42±11.6
2	1	0	0.87	4.35±11.4	0.87	4.35±11.4
2	1	1	0.86	4.33±11.3	0.87	4.33±11.4
2	1	2	0.87	4.34±11.4	0.86	4.35±11.4
2	1	3	0.88	4.34±11.5	0.85	4.36±11.6
2	1	4	0.87	4.35±11.5	0.86	4.37±11.6
2	1	5	0.87	4.33±11.4	0.86	4.39±11.7

Table C.8
Quadratic regression with t_1^* relative error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Rel. Error**	**Mean Rel. Error**	**Med. Rel. Error**	**Mean Rel. Error**
0	0	1	1.00	4.46±11.9	0.93	4.40±11.9
0	0	2	0.95	4.44±11.9	0.99	4.40±11.6
0	0	3	0.98	4.44±11.8	1.00	4.71±13.1
0	0	4	0.98	4.55±11.8	1.04	4.85±13.8
0	0	5	1.05	4.68±12.4	1.09	5.02±14.6
0	1	0	0.91	4.60±12.1	0.91	4.60±12.1
0	1	1	0.96	4.81±13.6	0.97	4.68±12.7
0	1	2	1.00	4.95±14.8	0.99	5.28±17.8
0	1	3	1.02	4.94±13.9	1.01	5.24±15.1
0	1	4	1.10	5.03±12.4	1.09	6.35±25.9
0	1	5	1.24	5.46±14.8	1.18	5.87±18.5
1	0	0	0.94	4.56±12.8	0.94	5.20±17.4
1	0	1	0.98	4.43±11.6	1.03	5.60±19.2
1	0	2	0.97	4.47±11.9	1.06	6.05±22.3
1	0	3	1.01	4.56±11.9	1.09	6.50±24.0
1	0	4	1.06	4.80±12.4	1.17	8.84±44.5
1	0	5	1.08	5.47±17.2	1.41	9.83±48.0
1	1	0	0.94	5.20±17.4	0.94	4.56±12.8
1	1	1	1.00	5.43±17.3	0.94	4.43±11.6
1	1	2	1.06	5.83±20.3	1.02	4.58±11.9
1	1	3	1.08	5.89±19.4	1.07	5.10±15.4
1	1	4	1.19	6.84±25.0	1.06	5.25±15.3
1	1	5	1.40	8.05±31.3	1.13	5.30±14.2
2	0	0	0.91	4.36±11.8	0.91	4.36±11.8
2	0	1	0.92	4.64±13.8	0.94	4.55±13.2
2	0	2	1.07	4.63±12.8	1.05	4.53±12.4
2	0	3	1.09	4.85±14.6	1.12	4.86±13.8
2	0	4	1.24	4.96±13.3	1.17	5.03±13.2
2	0	5	1.26	5.21±13.0	1.23	5.69±16.7
2	1	0	0.94	4.89±16.4	0.94	4.89±16.4
2	1	1	1.08	5.38±19.0	1.06	5.25±18.1
2	1	2	1.10	6.67±31.4	1.11	6.63±31.8
2	1	3	1.28	6.70±29.9	1.21	7.81±40.8
2	1	4	1.30	8.16±39.3	1.32	8.17±41.2
2	1	5	1.46	9.28±42.9	1.40	10.48±56.7

Table C.9
Gaussian Process with t_1^* relative error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Rel. Error**	**Mean Rel. Error**	**Med. Rel. Error**	**Mean Rel. Error**
0	0	1	1.02	4.59±12.3	1.01	4.70±12.6
0	0	2	1.01	4.54±12.1	1.02	4.70±12.8
0	0	3	1.02	4.58±12.3	1.01	4.70±12.9
0	0	4	1.00	4.60±12.4	1.01	4.70±12.9
0	0	5	1.00	4.60±12.5	1.01	4.70±12.9
0	1	0	0.91	4.48±11.8	0.91	4.48±11.8
0	1	1	0.91	4.50±12.0	0.89	4.42±11.7
0	1	2	0.95	4.51±12.0	0.92	4.48±11.8
0	1	3	0.95	4.54±12.1	0.94	4.56±12.2
0	1	4	0.96	4.57±12.2	0.95	4.57±12.3
0	1	5	0.96	4.58±12.3	0.96	4.59±12.3
1	0	0	0.99	4.57±12.2	0.93	4.39±11.4
1	0	1	1.01	4.53±11.8	0.92	4.36±11.3
1	0	2	1.01	4.52±11.9	0.93	4.39±11.3
1	0	3	1.04	4.55±12.0	0.95	4.44±11.6
1	0	4	1.04	4.58±12.1	0.95	4.46±11.7
1	0	5	1.03	4.58±12.2	0.95	4.46±11.7
1	1	0	0.93	4.39±11.4	0.99	4.57±12.2
1	1	1	0.90	4.39±11.4	1.00	4.58±12.1
1	1	2	0.94	4.39±11.4	1.00	4.55±11.9
1	1	3	0.96	4.41±11.5	1.04	4.64±12.4
1	1	4	0.99	4.44±11.5	1.03	4.64±12.4
1	1	5	1.00	4.48±11.8	1.03	4.61±12.2
2	0	0	0.99	4.40±11.1	0.99	4.40±11.1
2	0	1	1.01	4.44±11.5	1.01	4.41±11.4
2	0	2	1.02	4.44±11.5	1.00	4.43±11.4
2	0	3	1.02	4.46±11.6	1.01	4.49±11.7
2	0	4	1.03	4.49±11.7	0.99	4.50±11.8
2	0	5	1.02	4.50±11.8	0.98	4.51±11.8
2	1	0	0.89	4.37±11.2	0.89	4.37±11.2
2	1	1	0.90	4.37±11.3	0.92	4.36±11.2
2	1	2	0.92	4.38±11.3	0.92	4.37±11.3
2	1	3	0.92	4.39±11.4	0.92	4.41±11.4
2	1	4	0.94	4.41±11.4	0.92	4.43±11.6
2	1	5	0.93	4.41±11.4	0.93	4.43±11.5

Table C.10
SVR linear regression with t_1^* relative error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Rel. Error**	**Mean Rel. Error**	**Med. Rel. Error**	**Mean Rel. Error**
0	0	1	0.72	2.35±6.25	0.74	2.29±6.4
0	0	2	0.72	2.37±6.30	0.75	2.31±6.3
0	0	3	0.72	2.34±6.20	0.75	2.32±6.6
0	0	4	0.71	2.36±6.35	0.75	2.33±6.6
0	0	5	0.73	2.36±6.30	0.75	2.33±6.6
0	1	0	0.73	2.35±6.29	0.73	2.35±6.3
0	1	1	0.71	2.40±6.68	0.72	2.36±6.5
0	1	2	0.73	2.38±6.59	0.72	2.35±6.4
0	1	3	0.72	2.40±6.60	0.72	2.39±6.6
0	1	4	0.72	2.39±6.64	0.73	2.38±6.6
0	1	5	0.73	2.37±6.63	0.72	2.38±6.7
1	0	0	0.74	2.27±6.12	0.79	3.04±8.2
1	0	1	0.73	2.27±6.06	0.73	2.24±5.9
1	0	2	0.73	2.27±6.15	0.73	2.27±6.0
1	0	3	0.73	2.32±6.35	0.72	2.27±6.1
1	0	4	0.74	2.32±6.49	0.73	2.29±6.2
1	0	5	0.73	2.30±6.32	0.72	2.28±6.2
1	1	0	0.73	2.34±6.21	0.73	2.34±6.2
1	1	1	0.71	2.33±6.20	0.72	2.31±6.1
1	1	2	0.73	2.35±6.24	0.74	2.31±6.0
1	1	3	0.73	2.35±6.29	0.73	2.34±6.2
1	1	4	0.72	2.37±6.38	0.73	2.35±6.2
1	1	5	0.73	2.36±6.31	0.72	2.35±6.3
2	0	0	0.74	2.32±6.08	0.74	2.32±6.1
2	0	1	0.73	2.32±6.10	0.72	2.31±5.9
2	0	2	0.72	2.32±6.09	0.73	2.30±5.9
2	0	3	0.72	2.34±6.17	0.72	2.32±6.0
2	0	4	0.73	2.35±6.26	0.72	2.34±6.2
2	0	5	0.73	2.33±6.17	0.72	2.34±6.3
2	1	0	0.73	2.36±6.23	0.73	2.36±6.2
2	1	1	0.72	2.36±6.17	0.72	2.35±6.1
2	1	2	0.72	2.35±6.15	0.72	2.35±6.1
2	1	3	0.72	2.36±6.18	0.72	2.36±6.1
2	1	4	0.73	2.39±6.36	0.72	2.38±6.3
2	1	5	0.72	2.37±6.25	0.72	2.36±6.3

Table C.11
SVR with RBF kernel regression with t_1^* relative error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Rel. Error**	**Mean Rel. Error**	**Med. Rel. Error**	**Mean Rel. Error**
0	0	1	0.78	2.81±7.7	0.78	2.76±7.5
0	0	2	0.77	3.05±8.2	0.78	3.13±8.9
0	0	3	0.77	3.06±8.0	0.79	3.12±9.3
0	0	4	0.78	3.07±8.1	0.79	3.02±8.7
0	0	5	0.78	2.91±7.9	0.78	2.77±7.0
0	1	0	0.77	3.00±8.0	0.83	4.04±10.9
0	1	1	0.79	2.97±8.1	0.78	2.89±7.5
0	1	2	0.79	3.25±8.6	0.80	3.27±8.9
0	1	3	0.80	3.25±8.6	0.80	3.28±9.4
0	1	4	0.79	3.10±8.1	0.78	3.15±8.9
0	1	5	0.80	3.09±8.3	0.79	3.01±7.9
2	0	0	0.82	3.31±9.0	0.78	3.09±8.7
2	0	1	0.82	3.33±9.2	0.79	3.04±8.2
2	0	2	0.81	3.40±9.0	0.80	3.32±9.0
2	0	3	0.84	3.52±9.6	0.84	3.43±10.3
2	0	4	0.79	3.28±8.8	0.78	3.29±9.5
2	0	5	0.84	3.50±9.0	0.79	3.03±7.6
1	0	0	0.82	3.66±9.3	0.81	3.19±8.8
1	0	1	0.79	3.01±8.1	0.80	3.10±8.2
1	0	2	0.80	3.25±8.7	0.81	3.45±9.4
1	0	3	0.82	3.37±9.3	0.83	3.54±10.7
1	0	4	0.80	3.22±8.4	0.79	3.38±9.8
1	0	5	0.82	3.06±7.7	0.83	3.14±7.9
1	1	0	0.82	3.75±10.3	0.95	4.42±12.2
1	1	1	0.81	3.10±8.3	0.81	3.30±9.0
1	1	2	0.81	3.41±9.2	0.81	3.35±8.8
1	1	3	0.82	3.45±9.3	0.85	3.50±9.8
1	1	4	0.80	3.25±8.3	0.79	3.36±9.4
1	1	5	0.82	3.13±7.9	0.83	3.61±10.5
2	1	0	0.81	3.35±9.1	0.96	4.33±11.5
2	1	1	0.81	3.37±9.3	0.81	3.33±9.1
2	1	2	0.81	3.43±9.0	0.81	3.37±8.9
2	1	3	0.84	3.54±9.7	0.83	3.55±10.2
2	1	4	0.79	3.35±9.1	0.77	3.42±9.8
2	1	5	0.84	3.58±9.2	0.82	3.69±11.0

C.3 t_2^*

The following tables present the results of the regression evaluations for the complete DS1 data set using t_2^*. The order of the tables is linear regression, quadratic regression, Gaussian Process regression, SVR with a linear kernel and finally the SVR with the radial basis kernel for the absolute error. This order is then repeated for the relative error.

Table C.12
Linear regression with t_2^* absolute error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Abs. Error**	**Mean Abs. Error**	**Med. Abs. Error**	**Mean Abs. Error**
0	0	1	87.8	118.3±109.1	89.5	117.9±109.1
0	0	2	88.5	118.6±109.2	89.4	118.4±109.1
0	0	3	89.2	118.5±109.1	89.0	118.7±109.1
0	0	4	89.7	118.8±109.0	89.9	118.8±109.0
0	0	5	89.5	118.8±109.3	89.7	119.1±108.8
0	1	0	83.2	114.7±108.2	84.6	114.6±108.2
0	1	1	83.1	114.5±108.4	84.8	114.4±108.4
0	1	2	84.0	114.7±108.5	84.7	114.6±108.5
0	1	3	83.8	114.5±108.5	85.9	114.7±108.5
0	1	4	83.3	114.8±108.3	86.3	114.8±108.7
0	1	5	83.7	114.9±108.4	86.1	115.0±108.4
1	0	0	87.8	117.5±108.8	83.2	114.7±108.2
1	0	1	89.2	117.0±108.9	82.2	114.4±108.4
1	0	2	88.6	117.3±108.9	82.8	114.6±108.5
1	0	3	87.6	117.2±108.9	83.3	114.8±108.4
1	0	4	88.2	117.6±108.8	84.4	114.9±108.5
1	0	5	89.5	117.6±108.9	83.4	115.2±108.3
1	1	0	82.5	114.5±108.1	86.6	116.6±108.9
1	1	1	82.7	114.5±108.2	86.9	115.8±108.9
1	1	2	82.0	114.5±108.3	86.6	116.1±109.0
1	1	3	82.8	114.4±108.3	85.1	116.1±109.1
1	1	4	82.4	114.7±108.1	86.5	116.3±109.0
1	1	5	83.6	114.9±108.2	85.3	116.4±108.7
2	0	0	86.6	116.6±108.9	82.5	114.5±108.1
2	0	1	85.8	116.1±108.9	81.9	114.3±108.2
2	0	2	85.8	116.3±109.0	81.7	114.5±108.2
2	0	3	86.6	116.3±108.9	82.7	114.8±108.2
2	0	4	86.1	116.5±108.8	83.0	114.8±108.3
2	0	5	85.8	116.5±108.9	83.3	115.0±108.1
2	1	0	84.6	114.6±108.2	87.8	117.5±108.8
2	1	1	85.4	114.5±108.4	88.2	116.6±108.8
2	1	2	85.8	114.5±108.5	87.1	117.1±108.8
2	1	3	84.0	114.4±108.5	87.9	117.3±108.8
2	1	4	84.2	114.7±108.3	88.8	117.3±108.9
2	1	5	84.4	114.8±108.4	88.6	117.6±108.7

Table C.13
Quadratic regression with t_2^* absolute error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Abs. Error**	**Mean Abs. Error**	**Med. Abs. Error**	**Mean Abs. Error**
0	0	1	90.5	118.2±107.3	91.4	118.0±107.2
0	0	2	93.2	120.0±108.3	93.3	119.4±107.0
0	0	3	97.5	121.6±108.2	95.0	121.0±107.5
0	0	4	100.1	124.4±110.7	98.1	124.9±111.7
0	0	5	102.2	127.5±113.0	100.0	126.3±110.8
0	1	0	87.3	122.4±120.8	87.3	122.4±120.8
0	1	1	93.3	128.7±153.7	91.8	125.5±130.9
0	1	2	95.0	136.4±200.5	95.9	131.2±147.7
0	1	3	99.5	137.7±172.5	101.6	134.8±161.4
0	1	4	107.5	140.6±154.8	106.4	140.5±163.4
0	1	5	112.3	150.6±174.8	112.4	145.8±151.9
1	0	0	86.0	117.5±110.9	86.0	117.5±110.9
1	0	1	89.7	118.8±108.5	89.4	118.2±108.5
1	0	2	93.1	121.1±110.0	93.6	120.6±109.4
1	0	3	94.1	123.7±112.1	97.3	124.6±110.7
1	0	4	102.1	127.2±113.0	102.1	129.0±118
1	0	5	102.7	133.8±126.0	103.7	133.4±122.5
1	1	0	91.3	126.3±143.0	91.3	126.3±143.0
1	1	1	97.1	128.6±122.1	97.6	131.1±142.3
1	1	2	101.0	141.3±204.9	101.0	139.6±194.3
1	1	3	105.2	139.3±132.0	104.5	143.6±169.0
1	1	4	112.7	146.4±142.2	111.0	159.1±258.3
1	1	5	121.8	186.8±422.2	119.9	175.8±377.8
2	0	0	86.0	117.6±111.6	86.0	117.6±111.6
2	0	1	93.3	128.7±153.7	88.8	118.4±110.7
2	0	2	95.0	136.4±200.5	92.1	122.0±111.9
2	0	3	99.5	137.7±172.5	94.9	125.1±113.4
2	0	4	107.5	140.6±154.8	100.3	133.0±120.1
2	0	5	112.3	150.6±174.8	107.0	139.6±126.6
2	1	0	89.5	137.2±257.5	89.5	137.2±257.5
2	1	1	96.5	137.9±209.4	96.9	138.2±215.7
2	1	2	99.5	150.8±306.6	97.1	137.5±163.5
2	1	3	103.1	147.7±204.2	106.0	147.3±170.5
2	1	4	111.2	153.1±174.4	116.2	171.8±313.6
2	1	5	127.6	205.3±487.5	124.2	197.0±417.0

Table C.14
Gaussian Process regression with t_2^* absolute error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Abs. Error**	**Mean Abs. Error**	**Med. Abs. Error**	**Mean Abs. Error**
0	0	1	81.62	115.90±107.8	82.54	116.30±108.2
0	0	2	80.88	115.50±107.6	82.35	116.40±107.9
0	0	3	81.24	115.80±107.6	82.34	116.50±107.7
0	0	4	81.75	115.90±107.7	82.12	116.40±107.8
0	0	5	81.88	115.90±107.6	82.59	116.40±107.9
0	1	0	82.63	114.90±108.1	82.63	114.90±108.1
0	1	1	82.37	115.00±108.1	83.86	114.70±108.3
0	1	2	81.79	115.10±108.0	82.16	115.00±108.0
0	1	3	81.90	115.30±107.9	81.83	115.40±107.9
0	1	4	81.68	115.20±107.8	81.68	115.50±107.9
0	1	5	81.04	115.20±107.8	81.95	115.70±108.0
1	0	0	84.61	115.40±109.3	84.61	115.40±109.3
1	0	1	84.71	115.30±108.7	83.78	115.70±108.9
1	0	2	83.43	115.40±108.6	83.49	115.70±108.8
1	0	3	83.22	115.40±108.4	82.60	115.90±108.6
1	0	4	83.56	115.30±108.3	82.70	115.70±108.5
1	0	5	82.19	115.20±108.1	82.54	116.00±108.6
1	1	0	82.73	114.30±108.0	82.73	114.30±108.0
1	1	1	83.10	114.50±108.4	82.25	114.20±108.3
1	1	2	82.62	114.50±108.6	82.43	114.60±108.4
1	1	3	82.92	114.80±108.4	82.76	114.90±108.4
1	1	4	82.11	114.70±108.1	82.40	115.00±108.3
1	1	5	81.31	114.80±108.0	82.87	115.10±108.3
2	0	0	84.01	115.70±109.6	84.01	115.70±109.6
2	0	1	83.37	114.80±109.1	84.25	114.70±109.1
2	0	2	83.76	114.90±108.9	84.17	115.00±108.9
2	0	3	83.58	115.10±108.7	83.38	115.30±108.7
2	0	4	83.31	115.20±108.6	84.12	115.30±108.7
2	0	5	82.80	115.30±108.5	84.37	115.50±108.7
2	1	0	82.65	114.10±108.7	82.65	114.10±108.7
2	1	1	82.48	113.80±108.7	82.56	113.80±108.6
2	1	2	81.87	113.90±108.6	82.05	114.10±108.6
2	1	3	82.16	114.10±108.5	82.25	114.30±108.5
2	1	4	81.52	114.20±108.5	82.21	114.40±108.5
2	1	5	81.42	114.30±108.4	83.05	114.40±108.4

Table C.15
SVR linear regression with t_2^* absolute error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Abs. Error**	**Mean Abs. Error**	**Med. Abs. Error**	**Mean Abs. Error**
0	0	1	48.1	103.2±138.7	45.1	104.1±143.2
0	0	2	49.4	103.4±137.8	45.8	104.5±143.2
0	0	3	48.9	103.3±136.4	47.0	104.1±141.9
0	0	4	47.2	103.6±138.7	45.6	104.3±143.4
0	0	5	47.7	103.6±138.2	46.3	104.5±143.5
0	1	0	48.1	103.6±139.5	48.4	103.0±137.5
0	1	1	45.8	103.3±140.1	46.2	103.2±139.7
0	1	2	48.2	103.5±138.8	48.5	103.4±139.4
0	1	3	49.2	103.2±137.0	49.9	103.3±137.4
0	1	4	46.7	103.7±139.6	47.2	103.6±139.6
0	1	5	46.6	103.6±139.5	47.2	103.5±139.4
1	0	0	47.2	104.4±142.4	47.2	104.4±142.4
1	0	1	46.7	103.9±141.5	46.1	103.7±141.2
1	0	2	47.9	104.0±141.1	47.5	104.0±141.1
1	0	3	48.9	103.8±139.9	48.7	103.6±139.9
1	0	4	47.5	104.2±141.7	46.4	103.8±141.7
1	0	5	48.1	104.4±141.6	47.8	104.0±141.1
1	1	0	48.4	103.7±138.6	48.4	103.7±138.6
1	1	1	48.1	103.2±138.6	47.1	103.2±138.5
1	1	2	48.9	103.4±137.8	47.9	103.4±138.2
1	1	3	48.9	103.4±136.3	50.2	103.3±136.1
1	1	4	47.5	103.7±138.6	47.7	103.4±138.7
1	1	5	47.7	103.6±138.3	48.5	103.5±138.3
2	0	0	46.2	103.8±140.6	46.9	103.3±138.9
2	0	1	47.5	103.3±140.0	47.4	103.2±139.4
2	0	2	48.1	103.6±139.7	46.8	103.4±139.6
2	0	3	48.9	103.3±137.8	49.1	103.3±137.7
2	0	4	46.3	103.7±140.0	47.3	103.4±139.9
2	0	5	46.4	103.8±139.9	47.7	103.3±139.8
2	1	0	48.4	103.7±138.2	48.4	103.7±138.2
2	1	1	47.9	103.1±138.3	48.5	103.1±138.0
2	1	2	49.1	103.2±137.7	49.2	103.3±137.7
2	1	3	49.9	103.2±136.3	49.8	103.3±136.2
2	1	4	48.1	103.5±138.2	47.3	103.3±138.4
2	1	5	48.7	103.5±137.9	47.6	103.3±138.1

Table C.16
SVR with RBF kernel regression with t_2^* absolute error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Abs. Error**	**Mean Abs. Error**	**Med. Abs. Error**	**Mean Abs. Error**
0	0	1	58.3	109.2±133.7	57.2	108.5±135.1
0	0	2	62.3	111.3±133.4	61.0	111.1±134.9
0	0	3	52.8	106.6±136.7	52.7	106.5±137.4
0	0	4	62.6	112.2±133.4	59.7	111.4±135.1
0	0	5	61.2	110.8±134.4	60.0	111.0±135.6
0	1	0	59.9	109.7±134.0	62.4	110.5±131.7
0	1	1	61.1	110.2±132.3	60.7	110.1±133.2
0	1	2	65.5	113.2±132.1	64.9	113.2±133.5
0	1	3	55.9	107.7±136.8	57.3	107.5±137.0
0	1	4	64.7	113.1±132.4	65.1	113.3±133.5
0	1	5	64.2	112.0±131.5	64.2	112.1±132.9
1	0	0	61.5	110.7±134.7	61.5	110.7±134.7
1	0	1	62.1	109.6±133.2	61.2	109.1±133.1
1	0	2	65.6	114.4±133.0	64.6	113.0±133.7
1	0	3	55.7	108.0±137.4	55.3	107.7±137.9
1	0	4	64.2	113.0±132.6	64.1	112.5±133.3
1	0	5	63.0	111.7±133.4	63.2	111.9±134.1
1	1	0	64.2	112.1±132.5	43.4	69.7±90.8
1	1	1	63.3	110.6±133.2	62.3	110.2±133.3
1	1	2	66.3	113.8±132.1	66.1	113.2±132.4
1	1	3	56.9	108.0±136.4	56.8	107.8±136.5
1	1	4	66.1	112.8±131.7	65.2	113.0±132.0
1	1	5	64.4	112.6±132.3	64.3	112.8±132.6
2	0	0	64.0	111.5±132.8	64.0	111.5±132.8
2	0	1	63.4	110.8±132.5	63.3	110.8±133.0
2	0	2	65.6	113.2±131.5	66.2	113.4±132.0
2	0	3	57.5	108.1±136.0	57.9	108.6±136.1
2	0	4	64.8	112.1±132.5	64.7	112.6±133.0
2	0	5	64.0	111.2±131.3	64.7	111.8±132.1
2	1	0	64.6	111.4±132.3	64.6	111.4±132.3
2	1	1	62.7	110.9±132.3	62.5	111.0±132.6
2	1	2	65.8	113.0±131.3	66.4	113.5±131.6
2	1	3	57.2	107.9±135.4	57.0	108.3±135.4
2	1	4	65.6	111.9±131.7	65.2	112.2±132.1
2	1	5	64.9	111.5±130.7	65.5	112.1±131.3

Table C.17
Linear regression t_2^* relative error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Rel. Error**	**Mean Rel. Error**	**Med. Rel. Error**	**Mean Rel. Error**
0	0	1	1.04	5.26±14.9	1.03	5.23±14.9
0	0	2	1.04	5.22±14.6	1.04	5.20±14.7
0	0	3	1.06	5.17±14.8	1.07	5.16±14.6
0	0	4	1.05	5.16±14.8	1.02	5.17±14.7
0	0	5	1.03	5.19±14.9	1.06	5.24±14.9
0	1	0	0.93	5.11±14.9	0.88	5.04±14.8
0	1	1	0.89	5.12±15.0	0.89	5.04±14.8
0	1	2	0.89	5.06±14.7	0.88	5.05±14.8
0	1	3	0.90	5.04±15.0	0.90	5.03±14.7
0	1	4	0.90	5.05±14.9	0.92	5.05±14.8
0	1	5	0.90	5.07±15.0	0.93	5.09±14.9
1	0	0	1.04	5.02±13.9	0.93	5.11±14.9
1	0	1	1.02	4.99±14.0	0.89	5.11±15.0
1	0	2	1.05	4.95±13.6	0.91	5.11±15.0
1	0	3	1.06	4.92±13.8	0.92	5.09±14.9
1	0	4	1.05	4.92±13.8	0.93	5.07±14.9
1	0	5	1.00	4.94±13.9	0.92	5.13±15.1
1	1	0	0.92	5.00±14.5	1.02	5.00±13.9
1	1	1	0.88	4.98±14.5	0.98	4.97±14.0
1	1	2	0.92	4.92±14.1	1.01	4.98±14.1
1	1	3	0.89	4.90±14.4	1.02	4.95±14.0
1	1	4	0.89	4.92±14.3	1.00	5.00±14.2
1	1	5	0.91	4.93±14.4	1.03	5.03±14.3
2	1	0	0.88	5.04±14.8	0.92	5.00±14.5
2	1	1	0.88	5.04±14.8	0.90	4.95±14.4
2	1	2	0.88	4.97±14.4	0.92	4.96±14.4
2	1	3	0.87	4.99±14.7	0.91	4.95±14.3
2	1	4	0.87	4.99±14.7	0.92	4.94±14.3
2	1	5	0.86	5.00±14.8	0.93	5.00±14.5
2	0	0	1.02	5.00±13.9	1.04	5.01±13.9
2	0	1	1.00	4.98±14.0	1.02	4.93±13.8
2	0	2	1.01	4.94±13.7	1.03	4.92±13.8
2	0	3	1.01	4.92±14.0	1.06	4.90±13.7
2	0	4	1.05	4.92±13.9	1.02	4.90±13.7
2	0	5	1.06	4.93±14.0	1.00	4.97±14.0

Table C.18
Quadratic regression with t_2^* relative error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Rel. Error**	**Mean Rel. Error**	**Med. Rel. Error**	**Mean Rel. Error**
0	0	1	1.03	5.05±14.6	1.03	5.18±15.3
0	0	2	1.06	5.08±15.1	1.10	5.12±14.8
0	0	3	1.12	5.26±16.4	1.06	5.47±17.8
0	0	4	1.10	5.40±16.0	1.16	6.28±24.5
0	0	5	1.15	5.76±18.7	1.14	6.18±22.9
0	1	0	0.99	5.69±17.3	0.98	5.69±17.3
0	1	1	1.02	5.58±16.3	1.01	5.55±16.3
0	1	2	1.11	5.50±16.1	1.16	5.78±17.2
0	1	3	1.23	6.40±21.7	1.13	5.88±17.7
0	1	4	1.24	6.79±22.4	1.22	6.22±19.2
0	1	5	1.39	7.33±24.8	1.41	6.87±21.9
1	0	0	1.01	5.35±16.1	1.01	5.35±16.1
1	0	1	1.05	5.21±15.3	1.05	5.14±15.0
1	0	2	1.05	5.12±14.7	1.07	5.12±14.6
1	0	3	1.13	5.40±16.9	1.20	5.81±18.0
1	0	4	1.15	5.87±18.8	1.15	6.84±26.1
1	0	5	1.17	6.34±21.1	1.22	7.39±29.4
1	1	0	0.96	6.00±19.7	0.96	6.00±19.7
1	1	1	1.08	6.23±19.7	1.15	6.11±18.7
1	1	2	1.09	6.13±19.1	1.23	5.95±17.3
1	1	3	1.28	6.48±21.6	1.25	7.28±25.3
1	1	4	1.37	6.93±21.8	1.40	7.89±28.9
1	1	5	1.56	8.36±29.0	1.46	8.18±31.3
2	0	0	1.00	5.42±18.3	1.00	5.42±18.3
2	0	1	1.02	5.58±16.3	0.98	5.35±16.6
2	0	2	1.11	5.50±16.1	1.12	5.78±20.6
2	0	3	1.23	6.40±21.7	1.13	5.88±20.8
2	0	4	1.24	6.79±22.4	1.17	6.59±23.1
2	0	5	1.39	7.33±24.8	1.31	7.02±25.3
2	1	0	1.04	5.84±21.1	1.01	5.84±21.1
2	1	1	1.11	5.76±18.2	1.12	5.92±19.5
2	1	2	1.17	5.69±17.6	1.25	6.01±19.7
2	1	3	1.28	6.28±19.7	1.33	6.83±24.6
2	1	4	1.41	6.82±20.5	1.44	7.32±24.2
2	1	5	1.70	9.51±34.2	1.63	9.46±32.5

Table C.19
Gaussian Process regression with t_2^* relative error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Rel. Error**	**Mean Rel. Error**	**Med. Rel. Error**	**Mean Rel. Error**
0	0	1	1.03	5.20±14.5	1.06	5.28±14.8
0	0	2	1.07	5.16±14.4	1.09	5.20±14.4
0	0	3	1.08	5.10±14.1	1.10	5.21±14.5
0	0	4	1.06	5.12±14.1	1.11	5.21±14.5
0	0	5	1.08	5.13±14.2	1.11	5.23±14.6
0	1	0	0.95	5.17±14.8	0.95	5.17±14.8
0	1	1	0.97	5.17±14.7	0.99	5.14±14.8
0	1	2	0.99	5.18±14.7	1.04	5.14±14.6
0	1	3	0.98	5.15±14.6	1.05	5.15±14.6
0	1	4	0.99	5.15±14.6	1.05	5.18±14.7
0	1	5	0.99	5.17±14.7	1.02	5.21±14.8
1	0	0	1.08	5.14±13.8	1.08	5.14±13.8
1	0	1	1.05	5.10±14.0	1.08	5.11±13.9
1	0	2	1.05	5.10±13.9	1.05	5.07±13.6
1	0	3	1.07	5.05±13.7	1.03	5.08±13.6
1	0	4	1.04	5.06±13.7	1.05	5.09±13.9
1	0	5	1.06	5.05±13.7	1.07	5.11±13.7
1	1	0	0.98	4.94±14.0	0.98	4.94±14.0
1	1	1	0.98	4.97±13.9	0.98	4.91±13.8
1	1	2	0.99	4.94±13.8	1.02	4.94±13.9
1	1	3	1.00	4.95±13.8	1.01	4.94±13.8
1	1	4	1.02	5.03±14.0	1.01	4.99±14.0
1	1	5	1.05	5.04±14.0	1.01	4.98±13.9
2	0	0	1.07	5.13±14.1	1.09	5.13±14.1
2	0	1	1.06	5.08±14.1	1.06	5.08±14.2
2	0	2	1.07	5.09±14.1	1.07	5.07±14.1
2	0	3	1.08	5.06±13.9	1.07	5.07±14.1
2	0	4	1.09	5.08±14.0	1.08	5.11±14.1
2	0	5	1.08	5.10±14.1	1.09	5.13±14.2
2	1	0	1.01	5.06±14.3	1.01	5.06±14.3
2	1	1	1.00	5.02±14.2	1.02	5.02±14.2
2	1	2	1.01	5.01±14.1	1.03	5.03±14.2
2	1	3	1.01	5.01±14.1	1.03	5.04±14.2
2	1	4	1.02	5.04±14.2	1.03	5.06±14.3
2	1	5	1.02	5.06±14.2	1.01	5.05±14.3

Table C.20
SVR with linear kernel regression with t_2^* relative error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Rel. Error**	**Mean Rel. Error**	**Med. Rel. Error**	**Mean Rel. Error**
0	0	1	0.73	2.55±7.5	0.74	2.49±7.1
0	0	2	0.74	2.57±7.3	0.75	2.51±7.1
0	0	3	0.73	2.59±7.6	0.74	2.49±7.2
0	0	4	0.74	2.55±7.4	0.74	2.51±7.2
0	0	5	0.74	2.56±7.4	0.75	2.52±7.1
0	1	0	0.73	2.52±7.1	0.72	2.60±7.7
0	1	1	0.74	2.54±7.5	0.73	2.56±7.5
0	1	2	0.73	2.61±7.6	0.73	2.57±7.5
0	1	3	0.73	2.61±7.8	0.74	2.61±7.8
0	1	4	0.74	2.59±7.6	0.73	2.63±7.8
0	1	5	0.74	2.59±7.6	0.74	2.61±7.6
1	0	0	0.74	2.45±6.6	0.74	2.45±6.6
1	0	1	0.74	2.48±7.0	0.74	2.47±6.9
1	0	2	0.75	2.49±6.9	0.75	2.45±6.7
1	0	3	0.74	2.48±7.0	0.74	2.42±6.8
1	0	4	0.75	2.48±6.9	0.74	2.42±6.7
1	0	5	0.75	2.49±6.9	0.75	2.49±6.8
1	1	0	0.73	2.52±7.0	0.73	2.52±7.0
1	1	1	0.74	2.55±7.4	0.73	2.52±7.3
1	1	2	0.73	2.57±7.3	0.73	2.52±7.2
1	1	3	0.73	2.60±7.6	0.73	2.59±7.6
1	1	4	0.75	2.55±7.4	0.73	2.51±7.2
1	1	5	0.75	2.57±7.4	0.73	2.56±7.3
2	0	0	0.73	2.46±6.7	0.73	2.54±7.2
2	0	1	0.72	2.47±7.0	0.71	2.49±7.1
2	0	2	0.74	2.47±6.9	0.73	2.47±6.9
2	0	3	0.73	2.50±7.1	0.73	2.50±7.1
2	0	4	0.74	2.48±7.0	0.73	2.48±7.0
2	0	5	0.74	2.52±7.1	0.72	2.48±7.0
2	1	0	0.74	2.52±7.1	0.74	2.52±7.1
2	1	1	0.73	2.54±7.4	0.73	2.55±7.5
2	1	2	0.73	2.54±7.3	0.72	2.56±7.3
2	1	3	0.73	2.56±7.5	0.72	2.62±7.7
2	1	4	0.73	2.54±7.4	0.72	2.56±7.4
2	1	5	0.73	2.59±7.4	0.72	2.58±7.4

Table C.21
SVR regression with the RBF kernel using t_2^* relative error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Rel. Error**	**Mean Rel. Error**	**Med. Rel. Error**	**Mean Rel. Error**
0	0	1	0.79	3.45±9.6	0.79	3.45±10.0
0	0	2	0.83	3.57±9.7	0.81	3.71±10.8
0	0	3	0.79	3.07±9.0	0.79	3.00±8.7
0	0	4	0.83	3.55±9.6	0.82	3.55±10.0
0	0	5	0.80	3.63±10.4	0.81	3.55±10.1
0	1	0	0.80	3.53±10.3	0.82	3.72±10.4
0	1	1	0.82	3.66±10.4	0.81	3.53±9.8
0	1	2	0.87	3.81±10.4	0.88	3.95±11.2
0	1	3	0.80	3.21±9.5	0.78	3.16±9.2
0	1	4	0.85	3.76±10.4	0.86	3.77±10.3
0	1	5	0.84	3.80±10.7	0.84	3.74±10.3
1	0	0	0.82	3.49±9.3	0.82	3.49±9.3
1	0	1	0.83	3.80±11.4	0.81	3.67±10.6
1	0	2	0.92	4.04±11.2	0.88	4.01±11.4
1	0	3	0.79	3.06±8.5	0.78	3.09±8.6
1	0	4	0.87	3.77±10.2	0.85	3.78±10.2
1	0	5	0.84	3.83±10.9	0.85	3.85±11.2
1	1	0	0.89	3.94±11.2	0.78	3.37±8.8
1	1	1	0.84	3.86±11.6	0.84	3.78±11.2
1	1	2	0.94	4.06±11.5	0.91	4.10±11.7
1	1	3	0.80	3.16±8.7	0.79	3.20±8.9
1	1	4	0.91	3.84±10.4	0.89	3.86±10.4
1	1	5	0.87	3.90±11.0	0.86	3.97±11.4
2	0	0	0.87	3.93±11.1	0.87	3.93±11.1
2	0	1	0.82	3.98±12.2	0.81	3.94±11.9
2	0	2	0.92	4.10±11.5	0.92	4.12±11.7
2	0	3	0.79	3.21±9.0	0.79	3.18±8.9
2	0	4	0.87	3.86±10.7	0.87	3.91±10.9
2	0	5	0.85	3.92±11.3	0.84	3.98±11.6
2	1	0	0.87	3.96±11.2	0.87	3.96±11.2
2	1	1	0.83	3.93±12.0	0.82	3.89±11.4
2	1	2	0.91	4.09±11.4	0.91	4.15±11.6
2	1	3	0.81	3.26±9.1	0.79	3.21±9.0
2	1	4	0.87	3.86±10.7	0.87	3.90±10.7
2	1	5	0.86	3.99±11.4	0.86	4.03±11.6

C.4 Reduced Data Set t_1^*

The following tables present the results of the regression evaluations for the reduced DS1 data set using t_1^*. The order of the tables is linear regression, quadratic regression, Gaussian Process regression, SVR with a linear kernel and finally the SVR with the radial basis kernel for the absolute error. This order is then repeated for the relative error.

Table C.22
Linear regression with reduced data t_1^* absolute error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Abs. Error**	**Mean Abs. Error**	**Med. Abs. Error**	**Mean Abs. Error**
0	0	1	51.4	72.6±82.4	51.8	72.1±82.5
0	0	2	51.3	72.7±82.2	52.0	72.9±82.3
0	0	3	52.5	73.2±82.0	54.1	73.4±82.8
0	0	4	53.9	73.6±82.2	52.9	73.8±82.5
0	0	5	52.8	73.3±82.4	52.6	73.8±82.5
0	1	0	50.6	73.7±82.5	50.6	73.7±82.5
0	1	1	49.4	72.9±82.8	49.5	72.7±82.7
0	1	2	48.9	73.1±82.6	48.8	73.3±82.6
0	1	3	51.3	73.5±82.2	51.4	73.6±82.9
0	1	4	51.7	73.9±82.3	51.2	74.1±82.7
0	1	5	52.2	73.8±82.4	51.6	74.0±82.9
1	0	0	54.1	73.5±82.7	54.2	73.5±82.7
1	0	1	51.8	72.8±82.5	51.3	72.3±82.7
1	0	2	51.0	73.0±82.4	52.5	73.1±82.6
1	0	3	52.7	73.5±82.2	52.9	73.6±83.0
1	0	4	53.9	73.8±82.4	53.3	74.0±82.7
1	0	5	54.1	73.6±82.5	53.2	74.1±82.7
1	1	0	50.9	74.0±82.6	50.9	74.0±82.6
1	1	1	49.2	73.4±82.7	50.0	73.1±82.8
1	1	2	49.7	73.5±82.6	50.2	73.7±82.8
1	1	3	51.9	74.0±82.4	51.7	74.0±83.2
1	1	4	51.7	74.3±82.5	52.2	74.4±83.0
1	1	5	52.2	74.2±82.7	51.9	74.3±83.2
2	0	0	52.6	73.6±83.0	52.6	73.6±83.0
2	0	1	51.7	72.4±83.1	53.3	72.6±82.9
2	0	2	51.5	72.8±82.9	53.1	73.4±82.8
2	0	3	52.3	73.5±82.4	53.7	73.3±83.3
2	0	4	52.1	73.5±82.8	53.7	73.3±83.1
2	0	5	51.5	73.1±83.1	53.9	73.7±83.2
2	1	0	50.2	74.0±83.2	50.2	74.0±83.2
2	1	1	50.5	73.3±83.5	49.8	73.3±83.3
2	1	2	50.8	73.6±83.4	50.1	73.9±83.2
2	1	3	51.2	74.0±83.0	50.6	74.0±83.6
2	1	4	51.6	74.2±83.0	50.0	74.0±83.5
2	1	5	51.2	73.9±83.2	51.0	74.3±83.6

Table C.23
Quadratic regression with reduced data t_1^* absolute error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Abs. Error**	**Mean Abs. Error**	**Med. Abs. Error**	**Mean Abs. Error**
0	0	1	53.1	76.3±82.8	54.7	76.6±83.7
0	0	2	56.6	78.3±85.0	55.0	79.8±86.9
0	0	3	60.8	84.1±88.2	56.0	81.8±90.5
0	0	4	64.5	91.7±95.2	61.1	85.4±89.0
0	0	5	69.6	102.7±112.2	69.5	91.3±90.8
0	1	0	54.6	100.3±203.0	54.6	100.3±203.0
0	1	1	56.4	111.1±259.9	60.4	116.2±290.7
0	1	2	59.0	99.8±139.7	64.0	130.8±351.5
0	1	3	71.7	141.4±368.4	69.5	118.8±221.1
0	1	4	81.0	162.4±405.8	75.1	142.0±372.3
0	1	5	89.6	226.0±729.3	86.1	131.9±165.4
1	0	0	54.0	77.0±83.7	54.0	77.0±83.7
1	0	1	57.3	79.1±82.5	56.9	80.0±85.1
1	0	2	60.6	83.6±85.2	59.5	84.8±89.1
1	0	3	65.4	90.6±92.4	62.9	89.7±96.9
1	0	4	72.5	102.5±112.4	67.7	93.6±95.0
1	0	5	83.5	123.5±152.5	74.7	105.9±111.9
1	1	0	63.0	115.2±284.6	63.0	115.2±284.6
1	1	1	67.7	126.3±302.9	56.7	81.9±93.8
1	1	2	72.5	131.3±272.6	62.7	95.4±98.2
1	1	3	85.4	172.2±491.4	81.2	121.1±128.3
1	1	4	97.4	191.3±436.1	116.3	151.0±138.1
1	1	5	120.0	27.0±701.0	154.1	180.0±149.7
2	0	0	56.8	78.7±83.0	56.8	78.7±83.0
2	0	1	63.6	87.2±89.6	61.9	84.8±86.3
2	0	2	66.4	91.8±90.4	67.8	89.7±89.7
2	0	3	70.3	100.0±99.6	71.9	95.5±93.7
2	0	4	84.5	112.4±112.9	78.7	103.6±94.6
2	0	5	95.5	138.6±149.2	89.2	125.5±129.4
2	1	0	64.8	104.7±168.0	64.8	104.7±168.0
2	1	1	74.3	145.5±396.1	73.2	132.1±293.4
2	1	2	87.4	168.7±439.4	90.1	141.8±232.5
2	1	3	103.3	267.7±1017.0	95.4	151.8±222.4
2	1	4	129.9	376.0±1451.0	111.8	220.4±511.2
2	1	5	175.4	452.0±1401.0	153.4	345.5±836.6

Table C.24
Gaussian Process with reduced data t_1^* absolute error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Abs. Error**	**Mean Abs. Error**	**Med. Abs. Error**	**Mean Abs. Error**
0	0	1	53.9	73.2±81.3	54.2	72.6±81.0
0	0	2	54.0	73.3±81.3	54.8	73.0±80.8
0	0	3	54.9	73.5±81.2	55.4	73.6±81.0
0	0	4	55.1	73.7±81.3	55.8	73.3±80.6
0	0	5	55.5	73.6±81.1	55.9	73.5±81.0
0	1	0	55.0	73.5±81.7	55.0	73.5±81.7
0	1	1	53.9	72.8±81.3	52.7	72.5±81.1
0	1	2	53.4	72.8±81.1	53.1	72.7±80.9
0	1	3	54.1	73.1±81.0	54.0	73.2±81.0
0	1	4	54.1	73.4±81.1	55.0	73.1±80.7
0	1	5	54.4	73.4±81.1	54.6	73.3±81.0
1	0	0	54.6	73.0±81.3	54.6	73.0±81.3
1	0	1	54.6	73.1±80.9	54.6	73.1±80.9
1	0	2	54.2	73.3±81.1	54.2	73.3±81.1
1	0	3	54.3	73.4±81.1	54.3	73.4±81.1
1	0	4	55.4	73.6±81.2	55.4	73.6±81.2
1	0	5	55.9	73.5±81.0	55.9	73.5±81.0
1	1	0	55.2	73.2±81.2	55.2	73.2±81.2
1	1	1	53.6	72.9±81.1	53.6	72.9±81.1
1	1	2	54.0	72.9±81.0	54.0	72.9±81.0
1	1	3	54.3	73.1±81.0	54.3	73.1±81.0
1	1	4	53.7	73.4±81.1	53.7	73.4±81.1
1	1	5	55.3	73.7±81.4	55.3	73.7±81.4
2	0	0	54.9	73.5±81.4	54.9	73.5±81.4
2	0	1	53.3	73.1±81.2	53.7	72.9±81.2
2	0	2	54.7	73.2±81.4	55.2	73.1±81.2
2	0	3	54.7	73.4±81.3	56.5	73.7±81.2
2	0	4	55.3	73.7±81.4	56.4	73.3±80.9
2	0	5	56.1	73.6±81.3	56.2	73.5±81.2
2	1	0	52.8	73.2±81.0	52.8	73.2±81.0
2	1	1	53.9	73.2±81.3	52.1	72.5±81.1
2	1	2	54.0	73.2±81.3	52.9	72.7±81.0
2	1	3	54.9	73.5±81.2	54.3	73.2±81.0
2	1	4	55.1	73.7±81.3	54.7	73.0±80.8
2	1	5	55.5	73.6±81.1	54.4	73.2±81.1

Table C.25
SVR with linear kernel with reduced data t_1^* absolute error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Abs. Error**	**Mean Abs. Error**	**Med. Abs. Error**	**Mean Abs. Error**
0	0	1	32.5	64.4±96.3	50.5	103.3±135.2
0	0	2	32.2	64.0±96.5	50.5	103.4±135.1
0	0	3	32.9	64.6±96.6	49.7	103.1±135.3
0	0	4	32.6	64.9±96.3	50.3	103.3±135.5
0	0	5	33.4	64.6±96.1	45.0	103.2±135.9
0	1	0	31.8	64.4±97.0	31.8	64.4±97.0
0	1	1	31.6	64.4±96.4	48.6	103.3±136.6
0	1	2	31.5	64.3±96.6	48.2	103.4±136.6
0	1	3	31.9	64.7±96.9	49.2	103.4±136.3
0	1	4	33.1	65.1±96.8	48.6	103.4±136.9
0	1	5	33.3	64.6±96.5	47.9	103.4±137.6
1	0	0	31.6	64.4±97.3	31.6	64.4±97.3
1	0	1	48.1	103.8±139.0	31.4	63.8±96.7
1	0	2	47.9	103.8±139.5	31.0	63.9±97.4
1	0	3	49.4	104.0±138.9	32.1	64.3±97.4
1	0	4	48.9	104.2±139.5	32.2	64.5±97.1
1	0	5	48.1	104.3±140.1	31.9	64.2±97.2
1	1	0	31.7	64.5±96.9	31.7	64.5±96.9
1	1	1	50.2	103.0±135.8	31.8	64.3±96.7
1	1	2	51.1	103.3±135.5	32.4	64.2±97.0
1	1	3	50.7	103.1±135.4	33.3	64.5±97.1
1	1	4	50.7	103.6±135.8	34.0	64.9±96.8
1	1	5	50.2	103.5±136.2	32.7	64.5±97.1
2	0	0	31.4	64.0±97.1	31.4	64.0±97.1
2	0	1	32.5	63.8±96.2	49.9	103.9±137.5
2	0	2	32.2	63.7±96.5	49.4	103.8±137.7
2	0	3	32.4	64.0±96.7	49.1	103.8±137.2
2	0	4	32.2	64.4±96.5	49.1	103.7±137.6
2	0	5	31.5	64.2±96.4	49.3	103.7±137.6
2	1	0	31.5	64.3±96.7	31.5	64.3±96.7
2	1	1	31.4	63.9±96.6	48.1	104.4±141.4
2	1	2	30.7	63.9±96.8	48.4	104.5±141.1
2	1	3	31.4	64.3±97.1	47.6	104.6±141.6
2	1	4	31.6	64.6±97.0	47.6	104.5±141.7
2	1	5	32.3	64.4±96.9	47.8	104.6±142.0

Table C.26
SVR with RBF kernel with reduced data t_1^* absolute error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Abs. Error**	**Mean Abs. Error**	**Med. Abs. Error**	**Mean Abs. Error**
0	0	1	42.0	69.5±91.9	43.6	69.7±91.0
0	0	2	47.1	71.8±91.0	46.9	71.4±91.4
0	0	3	47.5	71.1±90.3	46.9	71.3±90.8
0	0	4	43.2	69.8±92.3	43.6	69.1±91.7
0	0	5	44.6	69.7±91.8	44.3	69.4±90.8
0	1	0	31.8	64.4±97.0	43.5	70.4±91.3
0	1	1	44.1	70.1±91.1	43.4	69.7±90.0
0	1	2	46.6	71.2±89.5	46.3	70.7±88.8
0	1	3	48.3	71.3±89.2	47.6	71.3±90.8
0	1	4	43.6	69.8±91.8	43.6	69.3±91.7
0	1	5	45.5	69.7±92.1	44.4	69.3±90.8
1	0	0	49.9	75.7±94.9	44.8	71.5±93.6
1	0	1	44.9	70.5±91.8	45.2	70.2±91.4
1	0	2	45.4	70.2±91.3	45.9	70.2±91.1
1	0	3	50.0	71.9±89.1	50.2	71.8±89.1
1	0	4	44.0	69.6±91.9	43.8	69.2±91.7
1	0	5	43.3	68.6±91.6	43.5	68.8±91.5
1	1	0	51.8	72.4±87.3	47.3	71.2±89.8
1	1	1	45.6	69.4±90.1	44.7	69.5±89.5
1	1	2	45.7	70.0±90.8	45.6	70.0±90.7
1	1	3	49.5	71.7±87.8	50.3	71.6±88.1
1	1	4	44.0	69.7±92.1	44.1	69.4±91.9
1	1	5	43.7	69.0±91.6	43.8	69.0±91.4
2	0	0	31.4	64.0±97.1	47.5	70.9±90.2
2	0	1	44.5	70.0±91.6	45.3	70.3±91.0
2	0	2	47.1	71.4±90.3	48.2	71.4±89.7
2	0	3	48.2	70.9±89.4	48.0	70.7±89.5
2	0	4	44.3	69.9±92.4	44.2	69.7±92.3
2	0	5	44.8	70.0±91.5	44.6	69.7±91.4
2	1	0	31.5	64.3±96.7	47.6	70.9±89.5
2	1	1	44.5	69.7±90.8	44.7	69.7±90.0
2	1	2	48.0	71.5±90.0	47.8	71.3±89.4
2	1	3	48.3	70.9±89.2	48.1	70.7±89.2
2	1	4	44.4	69.9±92.4	44.3	69.6±92.2
2	1	5	45.1	69.9±91.1	45.1	69.7±90.9

Table C.27
Linear regression with reduced data t_1^* relative error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Rel. Error**	**Mean Rel. Error**	**Med. Rel. Error**	**Mean Rel. Error**
0	0	1	0.83	3.58±9.4	0.84	3.52±9.2
0	0	2	0.83	3.56±9.2	0.81	3.54±9.1
0	0	3	0.84	3.55±9.0	0.83	3.63±9.4
0	0	4	0.83	3.58±9.0	0.81	3.67±9.6
0	0	5	0.82	3.59±9.2	0.82	3.68±9.7
0	1	0	0.84	3.77±10.5	0.84	3.77±10.5
0	1	1	0.79	3.71±10.5	0.79	3.66±10.4
0	1	2	0.79	3.67±10.3	0.80	3.67±10.2
0	1	3	0.80	3.67±10.2	0.81	3.74±10.5
0	1	4	0.81	3.70±10.2	0.83	3.78±10.8
0	1	5	0.80	3.70±10.3	0.82	3.80±11.0
1	0	0	0.82	3.66±9.3	0.82	3.66±9.3
1	0	1	0.82	3.57±9.3	0.82	3.52±9.2
1	0	2	0.82	3.55±9.2	0.82	3.55±9.1
1	0	3	0.84	3.55±9.0	0.82	3.62±9.3
1	0	4	0.82	3.59±9.0	0.81	3.67±9.5
1	0	5	0.82	3.61±9.2	0.81	3.68±9.7
1	1	0	0.82	3.75±10.3	0.82	3.75±10.3
1	1	1	0.80	3.66±10.2	0.80	3.63±10.2
1	1	2	0.80	3.63±10.0	0.81	3.67±10.1
1	1	3	0.81	3.65±10.0	0.81	3.71±10.3
1	1	4	0.81	3.68±10.0	0.82	3.74±10.5
1	1	5	0.80	3.68±10.1	0.81	3.77±10.8
2	0	0	0.85	3.57±9.0	0.85	3.57±9.0
2	0	1	0.80	3.47±9.0	0.82	3.46±9.1
2	0	2	0.83	3.45±8.9	0.82	3.47±8.9
2	0	3	0.81	3.44±8.7	0.82	3.55±9.2
2	0	4	0.81	3.47±8.7	0.81	3.58±9.7
2	0	5	0.80	3.48±8.9	0.82	3.59±9.7
2	1	0	0.82	3.78±10.7	0.82	3.78±10.7
2	1	1	0.79	3.66±10.5	0.79	3.67±10.6
2	1	2	0.80	3.65±10.3	0.79	3.68±10.5
2	1	3	0.77	3.65±10.3	0.80	3.74±10.8
2	1	4	0.77	3.68±10.3	0.79	3.75±11.1
2	1	5	0.78	3.67±10.4	0.80	3.78±11.4

Table C.28
Quadratic regression with reduced data t_1^* relative error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Rel. Error**	**Mean Rel. Error**	**Med. Rel. Error**	**Mean Rel. Error**
0	0	1	0.83	3.545±9.3	0.80	3.41±8.5
0	0	2	0.82	3.58±9.2	0.90	3.42±7.8
0	0	3	0.91	3.92±9.5	0.85	3.93±10.0
0	0	4	0.97	4.68±13.7	0.98	4.26±11.1
0	0	5	1.10	5.85±18.4	1.08	4.32±11.2
0	1	0	0.85	5.08±17.9	0.85	5.10±17.9
0	1	1	0.95	5.25±18.6	1.00	4.78±14.7
0	1	2	0.96	4.61±12.2	0.97	5.11±15.1
0	1	3	1.12	6.85±25.3	1.12	5.67±15.6
0	1	4	1.31	7.01±22.5	1.30	5.89±18.1
0	1	5	1.63	8.95±29.3	1.35	6.63±18.1
1	0	0	0.89	4.11±13.2	0.89	4.11±13.2
1	0	1	0.91	3.78±10.4	0.87	4.22±13.6
1	0	2	0.87	4.25±12.9	0.94	4.47±14.2
1	0	3	1.02	4.53±12.6	0.93	5.42±19.1
1	0	4	1.18	5.71±18.5	1.11	5.23±15.3
1	0	5	1.45	7.46±25.8	1.27	5.34±13.3
1	1	0	0.97	5.86±22.3	0.97	5.86±22.3
1	1	1	1.16	5.53±18.0	0.87	5.56±19.1
1	1	2	1.22	6.04±19.4	1.09	5.99±19.5
1	1	3	1.32	7.56±24.9	1.53	8.46±23.0
1	1	4	1.76	11.45±47.9	1.79	9.30±24.2
1	1	5	2.29	13.04±44.9	2.37	12.39±30.1
2	0	0	0.82	3.60±9.2	0.82	3.60±9.2
2	0	1	0.96	4.33±12.0	0.87	4.23±12.1
2	0	2	1.05	4.53±12.8	1.04	4.34±11.3
2	0	3	1.12	4.71±12.1	1.13	4.84±13.6
2	0	4	1.42	5.67±16.5	1.35	5.17±14.4
2	0	5	1.79	7.85±25.3	1.60	6.71±19.2
2	1	0	1.10	5.60±19.6	1.10	5.60±19.6
2	1	1	1.25	5.41±14.3	1.21	5.25±12.6
2	1	2	1.44	8.28±30.8	1.44	7.26±24.8
2	1	3	1.80	12.77±57.2	1.56	9.56±35.7
2	1	4	2.48	12.95±37.5	2.07	13.92±61.6
2	1	5	3.57	15.43±33.9	3.21	15.45±43.1

Table C.29
Gaussian Process with reduced data t_1^* relative error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Rel. Error**	**Mean Rel. Error**	**Med. Rel. Error**	**Mean Rel. Error**
0	0	1	0.84	3.68±9.3	0.85	3.62±9.2
0	0	2	0.85	3.66±9.2	0.88	3.62±9.1
0	0	3	0.85	3.69±9.3	0.87	3.72±9.5
0	0	4	0.84	3.74±9.5	0.87	3.72±9.6
0	0	5	0.85	3.75±9.7	0.87	3.73±9.5
0	1	0	0.84	3.69±9.4	0.84	3.69±9.4
0	1	1	0.85	3.66±9.4	0.85	3.62±9.4
0	1	2	0.85	3.63±9.4	0.84	3.60±9.2
0	1	3	0.84	3.65±9.4	0.85	3.70±9.5
0	1	4	0.86	3.70±9.4	0.85	3.72±9.7
0	1	5	0.85	3.73±9.6	0.86	3.71±9.5
1	0	0	0.83	3.74±9.7	0.83	3.74±9.7
1	0	1	0.85	3.70±9.5	0.85	3.70±9.5
1	0	2	0.84	3.68±9.4	0.84	3.68±9.4
1	0	3	0.86	3.71±9.5	0.86	3.71±9.5
1	0	4	0.84	3.75±9.6	0.84	3.75±9.6
1	0	5	0.84	3.77±9.8	0.84	3.77±9.8
1	1	0	0.82	3.76±9.8	0.82	3.76±9.8
1	1	1	0.85	3.68±9.5	0.85	3.68±9.5
1	1	2	0.85	3.65±9.4	0.85	3.65±9.4
1	1	3	0.85	3.68±9.4	0.85	3.68±9.4
1	1	4	0.84	3.72±9.5	0.84	3.72±9.5
1	1	5	0.86	3.73±9.5	0.86	3.73±9.5
2	0	0	0.85	3.74±9.6	0.85	3.74±9.6
2	0	1	0.86	3.68±9.4	0.87	3.63±9.2
2	0	2	0.88	3.64±9.2	0.89	3.63±9.2
2	0	3	0.89	3.67±9.3	0.86	3.73±9.5
2	0	4	0.86	3.72±9.5	0.86	3.72±9.6
2	0	5	0.86	3.75±9.6	0.86	3.74±9.6
2	1	0	0.82	3.71±9.5	0.82	3.70±9.5
2	1	1	0.84	3.68±9.3	0.85	3.62±9.4
2	1	2	0.85	3.66±9.2	0.85	3.61±9.3
2	1	3	0.85	3.69±9.3	0.85	3.69±9.5
2	1	4	0.84	3.74±9.5	0.85	3.70±9.6
2	1	5	0.85	3.75±9.7	0.85	3.70±9.6

Table C.30
SVR with linear kernel with reduced data t_1^* relative error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Rel. Error**	**Mean Rel. Error**	**Med. Rel. Error**	**Mean Rel. Error**
0	0	1	0.65	2.10±5.7	0.72	2.35±6.1
0	0	2	0.65	2.04±5.2	0.72	2.35±6.1
0	0	3	0.66	2.08±5.5	0.72	2.35±6.2
0	0	4	0.66	2.09±5.4	0.72	2.37±6.2
0	0	5	0.67	2.18±5.9	0.72	2.37±6.3
0	1	0	0.65	2.08±5.6	0.65	2.08±5.6
0	1	1	0.65	2.12±5.7	0.72	2.36±6.5
0	1	2	0.64	2.06±5.4	0.72	2.35±6.4
0	1	3	0.66	2.12±5.7	0.72	2.37±6.5
0	1	4	0.66	2.14±5.6	0.73	2.38±6.6
0	1	5	0.66	2.20±6.0	0.72	2.38±6.7
1	0	0	0.67	2.03±5.3	0.67	2.03±5.3
1	0	1	0.73	2.27±6.1	0.64	1.96±4.9
1	0	2	0.73	2.27±6.2	0.65	1.98±5.0
1	0	3	0.73	2.32±6.4	0.65	1.99±5.1
1	0	4	0.74	2.32±6.5	0.66	2.04±5.2
1	0	5	0.73	2.30±6.3	0.66	2.06±5.4
1	1	0	0.67	2.11±5.7	0.67	2.11±5.7
1	1	1	0.71	2.33±6.2	0.65	2.00±5.1
1	1	2	0.73	2.35±6.2	0.66	2.00±5.1
1	1	3	0.73	2.35±6.3	0.67	2.02±5.2
1	1	4	0.72	2.37±6.4	0.67	2.09±5.5
1	1	5	0.73	2.36±6.3	0.66	2.12±5.7
2	0	0	0.65	1.99±5.1	0.65	1.99±5.1
2	0	1	0.64	2.05±5.4	0.72	2.31±6.0
2	0	2	0.64	2.03±5.3	0.72	2.30±5.9
2	0	3	0.64	2.03±5.2	0.72	2.32±6.0
2	0	4	0.64	2.05±5.2	0.72	2.34±6.2
2	0	5	0.66	2.07±5.3	0.73	2.34±6.2
2	1	0	0.64	2.09±5.7	0.64	2.09±5.7
2	1	1	0.63	2.05±5.4	0.74	2.29±6.4
2	1	2	0.63	2.03±5.3	0.75	2.31±6.4
2	1	3	0.65	2.04±5.2	0.75	2.30±6.5
2	1	4	0.65	2.06±5.3	0.75	2.33±6.6
2	1	5	0.65	2.11±5.6	0.75	2.33±6.7

Table C.31
SVR with RBF kernel with reduced data t_1^* relative error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Rel. Error**	**Mean Rel. Error**	**Med. Rel. Error**	**Mean Rel. Error**
0	0	1	0.75	2.78±7.0	0.75	2.77±6.8
0	0	2	0.81	3.05±7.8	0.79	3.00±7.5
0	0	3	0.81	3.08±7.6	0.81	3.04±7.4
0	0	4	0.77	2.74±6.9	0.74	2.71±6.9
0	0	5	0.77	2.92±7.5	0.76	2.77±6.7
0	1	0	0.65	2.08±5.6	0.73	2.85±7.2
0	1	1	0.74	2.82±7.1	0.74	2.77±6.9
0	1	2	0.81	3.01±7.6	0.80	2.96±7.3
0	1	3	0.83	3.12±7.7	0.82	3.03±7.3
0	1	4	0.80	2.75±6.9	0.77	2.73±6.9
0	1	5	0.79	3.00±7.9	0.78	2.78±6.7
1	0	0	0.79	3.22±8.1	0.76	2.69±6.2
1	0	1	0.76	2.86±7.2	0.75	2.88±7.2
1	0	2	0.78	2.81±6.7	0.76	2.95±7.6
1	0	3	0.83	3.23±8.1	0.83	3.18±8.0
1	0	4	0.78	2.75±6.9	0.77	2.74±6.9
1	0	5	0.78	2.78±7.0	0.77	2.74±6.8
1	1	0	0.82	3.29±8.3	0.78	3.10±8.1
1	1	1	0.76	2.81±7.1	0.75	2.85±7.2
1	1	2	0.78	2.85±6.9	0.76	3.01±8.0
1	1	3	0.84	3.24±8.3	0.83	3.21±8.2
1	1	4	0.78	2.77±6.9	0.78	2.76±6.9
1	1	5	0.79	2.81±7.1	0.77	2.77±6.8
2	0	0	0.65	1.99±5.1	0.79	3.01±7.3
2	0	1	0.78	2.86±7.3	0.76	2.89±7.2
2	0	2	0.82	3.04±7.7	0.82	3.06±7.7
2	0	3	0.82	3.09±7.7	0.81	3.07±7.6
2	0	4	0.79	2.77±6.9	0.79	2.75±6.9
2	0	5	0.80	2.88±7.3	0.79	2.83±7.1
2	1	0	0.64	2.09±5.7	0.79	2.99±7.3
2	1	1	0.77	2.83±7.2	0.76	2.86±7.2
2	1	2	0.82	3.05±7.7	0.83	3.06±7.7
2	1	3	0.82	3.08±7.6	0.81	3.08±7.6
2	1	4	0.79	2.77±6.9	0.79	2.76±6.9
2	1	5	0.80	2.88±7.2	0.80	2.83±7.1

C.5 Reduced data set with t_2^*

The following tables present the results of the regression evaluations for the reduced DS1 data set using t_2^*. The order of the tables is linear regression, quadratic regression, Gaussian Process regression, SVR with a linear kernel and finally the SVR with the radial basis kernel for the absolute error. This order is then repeated for the relative error.

Table C.32
Linear regression with reduced data t_2^* absolute error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Abs. Error**	**Mean Abs. Error**	**Med. Abs. Error**	**Mean Abs. Error**
0	0	1	53.4	71.4±81.3	51.5	71.0±81.4
0	0	2	52.7	71.8±81.5	53.6	71.8±81.3
0	0	3	54.6	72.3±81.1	53.0	72.1±81.5
0	0	4	54.7	72.6±81.4	52.6	72.4±81.6
0	0	5	54.2	72.6±81.5	54.3	72.7±81.3
0	1	0	50.3	72.5±81.3	50.2	72.1±81.6
0	1	1	50.8	72.3±81.5	50.2	72.1±81.6
0	1	2	50.6	72.6±81.6	49.8	72.5±81.7
0	1	3	51.1	72.8±81.3	50.0	72.7±81.8
0	1	4	50.9	73.1±81.5	49.6	73.0±82.0
0	1	5	51.2	73.1±81.7	52.0	73.3±81.8
1	0	0	53.4	72.2±81.3	53.4	72.2±81.3
1	0	1	53.5	71.8±81.3	52.2	71.4±81.4
1	0	2	53.4	72.2±81.6	52.8	72.0±81.6
1	0	3	54.9	72.5±81.3	53.0	72.2±81.8
1	0	4	54.8	72.8±81.6	53.2	72.5±81.9
1	0	5	54.5	72.8±81.8	54.4	72.9±81.6
1	1	0	50.0	72.9±81.4	50.0	72.9±81.4
1	1	1	51.0	72.7±81.6	50.6	72.4±81.6
1	1	2	51.7	73.0±81.7	49.2	72.7±81.9
1	1	3	52.7	73.1±81.5	49.0	73.0±82.0
1	1	4	51.8	73.4±81.7	49.6	73.2±82.1
1	1	5	52.3	73.4±81.9	51.0	73.6±81.8
2	0	0	53.2	72.2±81.4	51.1	71.2±81.4
2	0	1	51.5	71.3±81.2	51.1	71.2±81.4
2	0	2	51.3	71.6±81.6	51.7	71.8±81.4
2	0	3	52.4	72.1±81.2	53.6	72.1±81.5
2	0	4	53.5	72.4±81.5	53.2	72.6±81.5
2	0	5	52.3	72.4±81.7	54.6	72.9±81.0
2	1	0	50.9	72.9±82.1	50.5	72.4±82.2
2	1	1	52.1	72.5±82.1	50.5	72.4±82.2
2	1	2	51.3	72.7±82.4	49.6	72.6±82.4
2	1	3	52.1	73.0±82.0	50.3	72.9±82.3
2	1	4	51.8	73.3±82.2	50.4	73.4±82.5
2	1	5	52.2	73.3±82.4	51.3	73.7±82.0

Table C.33
Quadratic regression with reduced data t_2^* absolute error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Abs. Error**	**Mean Abs. Error**	**Med. Abs. Error**	**Mean Abs. Error**
0	0	1	53.7	75.7±80.5	57.0	76.5±80.7
0	0	2	53.8	77.9±83.3	56.8	77.9±82.1
0	0	3	60.3	83.1±87	61.0	81.1±85.5
0	0	4	63.5	91.6±107.1	66.7	87.6±89.4
0	0	5	70.8	98.0±100.7	67.6	90.0±89.6
0	1	0	57.5	100.9±192.9	56.1	78.0±84.3
0	1	1	62.1	102.4±160.3	62.9	104.7±173.4
0	1	2	63.5	110.5±188.5	66.3	117.8±200.2
0	1	3	72.1	120.3±183.1	68.7	140.2±357.9
0	1	4	77.3	149.9±302.1	78.6	163.2±442.7
0	1	5	92.3	243.3±802.4	81.2	153.3±324.6
1	0	0	56.2	75.3±81.7	56.2	75.3±81.7
1	0	1	57.8	77.8±81.0	57.7	77.6±81.3
1	0	2	58.2	82.4±85.3	62.1	82.3±83.7
1	0	3	65.2	91.1±93.0	66.4	86.5±87.9
1	0	4	72.3	104.7±120.5	73.1	99.2±100.4
1	0	5	84.0	126.7±149.8	76.7	104.1±100.0
1	1	0	61.4	104.0±179.5	61.4	104.0±179.5
1	1	1	68.1	109.3±165.3	67.8	107.2±146.3
1	1	2	74.1	119.0±157.7	76.8	124.8±179.8
1	1	3	85.7	164.9±357.6	81.8	186.9±570.0
1	1	4	103.7	240.7±653.8	99.4	205.8±586.3
1	1	5	130.9	429.0±1708.0	109.8	252.0±778.7
2	0	0	56.1	78.0±84.3	56.1	78.0±84.3
2	0	1	59.7	83.0±86.3	60.8	83.5±86.0
2	0	2	64.8	89.9±92.1	66.3	87.0±87.7
2	0	3	71.7	102.8±108.6	74.3	97.7±98.0
2	0	4	88.0	122.9±130.1	82.4	115.0±117.9
2	0	5	100.3	151.3±163.4	96.2	134.1±144.4
2	1	0	65.5	105.8±157.2	65.5	105.8±157.2
2	1	1	73.7	133.4±307.9	78.7	123.3±204.2
2	1	2	88.7	154.9±315.6	83.4	139.0±213.0
2	1	3	102.7	161.1±261.3	95.2	180.8±433.2
2	1	4	131.5	230.2±351.8	116.9	261.0±854.4
2	1	5	171.2	470.1±1390.0	165.1	353.1±902.3

Table C.34
Gaussian Process with reduced data t_2^* absolute error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Abs. Error**	**Mean Abs. Error**	**Med. Abs. Error**	**Mean Abs. Error**
0	0	1	54.4	72.2±80.6	55.3	71.8±80.4
0	0	2	55.3	72.3±80.7	55.8	72.3±80.8
0	0	3	55.8	72.5±80.6	55.5	72.5±80.7
0	0	4	55.8	72.5±80.7	55.5	72.4±80.5
0	0	5	55.3	72.5±80.7	55.5	72.6±80.9
0	1	0	54.1	72.0±80.7	54.1	72.0±80.7
0	1	1	53.4	71.9±80.3	53.1	71.6±80.3
0	1	2	53.2	71.9±80.3	54.1	72.1±80.4
0	1	3	53.8	72.1±80.2	54.6	72.2±81.0
0	1	4	53.8	72.1±80.2	54.6	72.3±80.3
0	1	5	54.9	72.3±80.4	54.8	72.4±80.4
1	0	0	54.3	71.9±80.9	54.3	71.9±80.9
1	0	1	54.1	72.0±80.4	54.1	71.5±80.3
1	0	2	54.8	72.2±80.6	55.0	72.2±80.6
1	0	3	55.7	72.4±80.5	55.2	72.5±80.7
1	0	4	55.2	72.4±80.6	54.6	72.3±80.5
1	0	5	54.8	72.3±80.6	55.2	73.7±81.7
1	1	0	54.3	71.9±80.6	54.3	71.9±80.6
1	1	1	53.3	71.9±80.2	52.6	71.5±80.2
1	1	2	53.3	72.0±80.3	54.7	73.2±81.4
1	1	3	54.4	73.3±81.1	55.1	73.4±81.4
1	1	4	54.5	73.2±81.2	54.8	73.4±81.3
1	1	5	54.6	73.3±81.7	54.6	73.5±81.4
2	0	0	55.2	72.5±81.6	55.2	72.5±81.6
2	0	1	53.8	72.2±80.8	54.7	72.0±80.9
2	0	2	54.2	72.2±80.8	55.9	72.4±80.9
2	0	3	54.3	72.5±80.8	56.3	72.6±80.9
2	0	4	55.6	72.5±80.9	56.2	72.5±80.8
2	0	5	55.6	72.7±81.1	56.0	72.5±80.9
2	1	0	55.1	72.1±81.0	55.1	72.1±81.0
2	1	1	53.4	71.8±80.6	53.3	71.7±80.7
2	1	2	53.5	73.1±81.8	54.3	72.0±80.7
2	1	3	54.6	73.8±81.9	54.5	72.4±81.1
2	1	4	55.3	73.8±81.9	55.2	72.5±80.8
2	1	5	54.8	73.5±81.9	55.4	72.5±80.7

Table C.35
SVR with linear kernel with reduced data t_2^* absolute error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Abs. Error**	**Mean Abs. Error**	**Med. Abs. Error**	**Mean Abs. Error**
0	0	1	30.6	62.9±95.9	30.2	62.9±96.0
0	0	2	30.7	63.3±96.0	29.9	63.4±96.3
0	0	3	31.0	63.6±96.0	30.0	63.4±96.0
0	0	4	30.0	63.4±96.2	30.2	63.2±96.1
0	0	5	29.5	63.4±96.6	30.0	63.3±96.4
0	1	0	30.0	63.5±96.0	30.0	63.5±96.0
0	1	1	31.5	63.3±95.6	30.9	63.2±95.5
0	1	2	31.7	63.9±95.8	30.8	63.7±95.9
0	1	3	30.8	63.7±95.8	30.4	63.7±95.9
0	1	4	29.9	63.7±95.8	30.2	63.5±95.9
0	1	5	29.6	63.5±96.3	29.9	63.5±96.0
1	0	0	30.6	63.2±96.1	30.6	63.1±96.1
1	0	1	31.5	63.3±95.6	31.4	62.7±95.9
1	0	2	31.7	63.9±95.8	30.8	63.1±96.3
1	0	3	30.8	63.7±95.8	30.8	63.1±96.0
1	0	4	29.9	63.7±95.8	30.5	63.1±96.2
1	0	5	29.6	63.5±96.3	30.5	63.0±96.4
1	1	0	30.4	63.3±96.0	30.4	63.3±96.0
1	1	1	31.6	63.2±95.7	32.1	63.1±95.7
1	1	2	31.6	63.6±95.9	30.9	63.5±96.0
1	1	3	31.1	63.6±95.9	30.8	63.5±95.8
1	1	4	30.8	63.4±96.0	30.6	63.3±96.0
1	1	5	30.6	63.5±96.2	30.8	63.3±96.2
2	0	0	30.1	63.3±96.1	30.1	63.3±96.1
2	0	1	31.2	62.8±95.0	47.1	103.3±139.6
2	0	2	31.6	63.1±95.2	47.5	103.4±139.6
2	0	3	31.3	63.2±94.8	49.1	103.4±137.9
2	0	4	31.2	63.2±95.2	46.9	103.3±140.2
2	0	5	30.8	63.2±95.9	47.5	103.3±139.8
2	1	0	30.8	63.5±95.8	30.8	63.5±95.8
2	1	1	31.2	62.8±95.1	32.0	63.2±95.5
2	1	2	31.6	63.1±95.1	31.4	63.5±95.5
2	1	3	31.0	63.2±94.9	31.0	63.4±95.2
2	1	4	31.2	63.1±95.1	30.6	63.4±95.5
2	1	5	30.9	63.2±95.8	31.2	63.5±95.8

Table C.36
SVR with RBF kernel with reduced data t_2^* absolute error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Abs. Error**	**Mean Abs. Error**	**Med. Abs. Error**	**Mean Abs. Error**
0	0	1	49.4	73.3±89.5	48.4	71.7±89.5
0	0	2	45.4	72.9±92.0	47.3	72.2±91.4
0	0	3	45.7	69.9±89.8	44.3	68.8±89.4
0	0	4	43.7	70.4±91.1	44.9	70.8±90.2
0	0	5	43.4	68.9±89.2	42.5	67.1±86.8
0	1	0	43.1	70.9±92.4	45.2	72.2±91.6
0	1	1	49.0	72.8±90.0	48.2	72.2±90.2
0	1	2	46.6	72.6±91.5	47.1	72.9±91.1
0	1	3	46.0	69.9±89.5	45.1	69.0±88.8
0	1	4	44.6	71.2±91.1	45.4	71.7±90.7
0	1	5	44.3	70.3±90.0	43.8	67.5±86.7
1	0	0	41.6	70.1±91.7	41.6	70.1±91.7
1	0	1	49.5	72.3±88.0	47.5	71.2±89.0
1	0	2	47.7	73.9±90.5	47.2	72.5±90.3
1	0	3	49.0	72.3±87.2	48.3	72.2±88.6
1	0	4	43.9	70.5±91.2	44.5	70.8±91.5
1	0	5	42.9	69.3±89.3	42.7	69.1±90.0
1	1	0	43.4	69.7±90.8	43.4	69.7±90.8
1	1	1	49.8	73.2±89.4	47.7	71.7±90.3
1	1	2	48.0	73.6±90.6	47.2	72.5±89.8
1	1	3	49.3	73.1±88.4	49.0	72.7±87.9
1	1	4	44.4	70.8±91.4	45.1	71.3±91.4
1	1	5	43.4	69.8±90.4	43.4	69.1±89.2
2	0	0	44.2	69.9±91.0	44.2	69.9±91.0
2	0	1	50.2	73.0±87.7	50.2	73.0±87.7
2	0	2	47.1	71.5±89.2	47.1	71.5±89.2
2	0	3	46.3	69.5±87.3	46.3	69.5±87.3
2	0	4	46.3	71.3±89.3	46.3	71.3±89.3
2	0	5	47.2	70.5±87.9	47.1	70.5±87.9
2	1	0	45.0	70.1±91.2	45.0	70.1±91.2
2	1	1	51.0	73.1±88.2	50.9	73.2±87.7
2	1	2	47.7	72.2±88.9	47.6	71.6±89.2
2	1	3	47.4	70.5±87.8	46.9	69.9±87.2
2	1	4	45.7	70.6±89.3	46.0	71.1±89.2
2	1	5	46.4	69.8±88.4	46.0	71.1±89.2

Table C.37
Linear regression with reduced data t_2^* relative error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Rel. Error**	**Mean Rel. Error**	**Med. Rel. Error**	**Mean Rel. Error**
0	0	1	0.84	3.90±10.6	0.88	3.84±10.4
0	0	2	0.89	3.87±10.3	0.88	3.76±9.7
0	0	3	0.89	3.79±9.5	0.88	3.80±9.7
0	0	4	0.90	3.85±9.8	0.89	3.78±9.7
0	0	5	0.87	3.85±9.8	0.88	3.91±10.3
0	1	0	0.82	3.87±10.3	0.80	3.85±10.6
0	1	1	0.80	3.90±10.8	0.80	3.85±10.6
0	1	2	0.85	3.89±10.6	0.82	3.80±10.2
0	1	3	0.85	3.84±10.1	0.83	3.83±10.3
0	1	4	0.83	3.88±10.4	0.83	3.84±10.3
0	1	5	0.83	3.86±10.3	0.84	3.93±10.7
1	0	0	0.91	3.89±10.0	0.91	3.89±10.0
1	0	1	0.83	3.93±10.7	0.89	3.85±10.3
1	0	2	0.91	3.91±10.3	0.90	3.80±9.8
1	0	3	0.88	3.87±9.9	0.90	3.84±9.8
1	0	4	0.90	3.93±10.2	0.89	3.81±9.8
1	0	5	0.87	3.92±10.1	0.86	3.97±10.5
1	1	0	0.84	3.91±10.4	0.83	3.91±10.4
1	1	1	0.80	3.90±10.7	0.80	3.84±10.3
1	1	2	0.85	3.89±10.5	0.83	3.80±10.1
1	1	3	0.84	3.87±10.2	0.81	3.84±10.2
1	1	4	0.83	3.93±10.5	0.83	3.82±10.2
1	1	5	0.83	3.91±10.4	0.85	3.95±10.6
2	0	0	0.92	3.70±9.3	0.87	3.71±9.9
2	0	1	0.87	3.74±10.2	0.87	3.71±9.9
2	0	2	0.88	3.72±9.9	0.88	3.65±9.3
2	0	3	0.92	3.64±9.3	0.89	3.69±9.5
2	0	4	0.89	3.70±9.5	0.92	3.74±9.7
2	0	5	0.88	3.70±9.5	0.87	3.83±10.2
2	1	0	0.85	3.82±10.2	0.82	3.81±10.4
2	1	1	0.81	3.85±10.7	0.82	3.81±10.4
2	1	2	0.84	3.83±10.5	0.84	3.78±10.2
2	1	3	0.84	3.80±10.2	0.83	3.82±10.4
2	1	4	0.86	3.85±10.5	0.84	3.86±10.5
2	1	5	0.85	3.83±10.4	0.83	3.95±11.0

Table C.38
SVR with RBF kernel regression with reduced data and t_2^* absolute error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Rel. Error**	**Mean Rel. Error**	**Med. Rel. Error**	**Mean Rel. Error**
0	0	1	0.96	4.06±11.1	0.96	4.06±11.1
0	0	2	0.96	4.22±12.0	0.94	4.09±11.5
0	0	3	0.96	5.12±17.3	1.00	4.89±16.7
0	0	4	1.01	5.34±16.8	1.02	5.03±15.3
0	0	5	1.17	5.39±15.3	1.08	5.60±19.7
0	1	0	0.90	4.98±14.0	0.89	3.73±9.2
0	1	1	1.15	4.92±13.3	1.06	4.99±13.5
0	1	2	1.09	5.43±16.5	1.18	6.28±19.9
0	1	3	1.25	7.93±30.7	1.28	7.27±25.6
0	1	4	1.37	9.12±35.0	1.34	8.56±31.0
0	1	5	1.70	11.56±43.1	1.52	8.22±28.2
1	0	0	0.91	4.2±12.1	0.91	4.2±12.1
1	0	1	0.93	4.04±10.6	0.98	3.58±8.5
1	0	2	0.99	4.45±12.5	1.03	4.03±10.2
1	0	3	1.05	5.38±17.3	1.12	5.12±16.4
1	0	4	1.14	5.51±15.4	1.18	5.61±16.1
1	0	5	1.49	8.34±31.7	1.20	6.60±20.8
1	1	0	1.01	5.80±19.0	1.01	5.80±19.0
1	1	1	1.20	6.49±21.4	1.18	5.94±18.3
1	1	2	1.35	6.97±23.9	1.26	7.63±25.5
1	1	3	1.32	10.49±43.2	1.43	9.20±32.4
1	1	4	1.84	15.57±69.0	1.69	10.39±36.7
1	1	5	2.35	22.78±105.4	2.30	12.16±43.4
2	0	0	0.89	3.73±9.2	0.89	3.73±9.2
2	0	1	0.96	4.70±15.7	0.98	4.87±16.6
2	0	2	1.03	5.35±18.2	0.99	4.85±14.9
2	0	3	1.16	7.16±29.2	1.16	5.31±16.0
2	0	4	1.44	8.10±31.2	1.41	7.59±26.4
2	0	5	1.76	11.68±47.9	1.68	10.49±44.1
2	1	0	1.09	5.80±17.5	1.09	5.80±17.5
2	1	1	1.33	6.61±20.9	1.33	5.81±16.3
2	1	2	1.51	7.62±22.7	1.50	7.34±21.8
2	1	3	1.70	9.58±32.6	1.70	8.99±29.2
2	1	4	2.19	12.09±33.3	2.13	12.37±44.5
2	1	5	3.31	23.93±88.8	3.34	24.21±99.9

Table C.39
Gaussian Process with reduced data t_2^* relative error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Rel. Error**	**Mean Rel. Error**	**Med. Rel. Error**	**Mean Rel. Error**
0	0	1	0.88	4.04±10.6	0.87	4.00±10.4
0	0	2	0.88	4.03±10.6	0.88	4.01±10.3
0	0	3	0.87	4.00±10.3	0.89	4.01±10.3
0	0	4	0.89	4.03±10.4	0.90	4.01±10.3
0	0	5	0.91	4.06±10.5	0.90	4.08±10.6
0	1	0	0.87	4.06±10.7	0.87	4.06±10.7
0	1	1	0.85	4.03±10.7	0.87	4.00±10.5
0	1	2	0.88	3.99±10.5	0.91	3.97±10.2
0	1	3	0.89	3.95±10.2	0.94	3.96±10.2
0	1	4	0.88	4.00±10.4	0.89	4.00±10.3
0	1	5	0.91	4.03±10.5	0.90	4.05±10.5
1	0	0	0.89	3.98±10.3	0.89	3.98±10.3
1	0	1	0.88	4.00±10.4	0.87	3.97±10.3
1	0	2	0.89	4.02±10.5	0.89	4.01±10.4
1	0	3	0.90	4.01±10.3	0.90	4.01±10.3
1	0	4	0.88	4.04±10.5	0.89	4.01±10.3
1	0	5	0.88	4.04±10.5	0.92	4.10±10.6
1	1	0	0.86	4.02±10.5	0.86	4.02±10.5
1	1	1	0.86	4.01±10.6	0.08	3.98±10.4
1	1	2	0.88	4.00±10.5	0.93	4.02±10.3
1	1	3	0.91	4.02±10.3	0.93	4.02±10.3
1	1	4	0.90	4.05±10.5	0.90	4.03±10.4
1	1	5	0.90	4.06±10.5	0.92	4.08±10.5
2	0	0	0.91	4.02±10.4	0.91	4.02±10.4
2	0	1	0.88	3.97±10.3	0.88	3.96±10.3
2	0	2	0.89	3.95±10.2	0.89	3.95±10.1
2	0	3	0.87	3.94±10.0	0.90	3.98±10.2
2	0	4	0.90	3.99±10.3	0.89	3.98±10.2
2	0	5	0.91	4.03±10.4	0.89	4.01±10.3
2	1	0	0.90	3.96±10.2	0.90	3.96±10.2
2	1	1	0.88	3.93±10.3	0.89	3.92±10.2
2	1	2	0.93	3.98±10.3	0.90	3.90±10.0
2	1	3	0.93	3.98±10.1	0.93	3.94±10.1
2	1	4	0.91	4.04±10.4	0.90	3.96±10.1
2	1	5	0.91	4.03±10.4	0.93	4.00±10.3

Table C.40
SVR with linear kernel and with reduced data t_2^* relative error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Rel. Error**	**Mean Rel. Error**	**Med. Rel. Error**	**Mean Rel. Error**
0	0	1	0.63	2.14±5.9	0.64	2.10±5.8
0	0	2	0.64	2.13±5.8	0.65	2.03±5.3
0	0	3	0.66	2.11±5.6	0.66	2.08±5.5
0	0	4	0.67	2.06±5.5	0.67	2.06±5.4
0	0	5	0.66	2.11±5.6	0.66	2.14±5.8
0	1	0	0.67	2.03±5.2	0.67	2.03±5.2
0	1	1	0.64	2.16±6.0	0.65	2.14±5.9
0	1	2	0.67	2.16±5.8	0.66	2.06±5.3
0	1	3	0.65	2.12±5.6	0.67	2.07±5.4
0	1	4	0.67	2.09±5.6	0.68	2.08±5.5
0	1	5	0.66	2.10±5.5	0.66	2.13±5.7
1	0	0	0.66	2.19±5.8	0.66	2.19±5.8
1	0	1	0.64	2.16±6.0	0.62	2.12±5.8
1	0	2	0.66	2.16±5.8	0.63	2.03±5.3
1	0	3	0.65	2.12±5.6	0.64	2.09±5.5
1	0	4	0.67	2.09±5.6	0.66	2.06±5.5
1	0	5	0.66	2.10±5.5	0.63	2.15±5.7
1	1	0	0.66	2.18±5.8	0.66	2.18±5.8
1	1	1	0.63	2.21±6.1	0.63	2.16±5.8
1	1	2	0.64	2.19±5.9	0.63	2.06±5.3
1	1	3	0.63	2.18±5.8	0.63	2.12±5.5
1	1	4	0.65	2.19±6.0	0.66	2.06±5.4
1	1	5	0.64	2.20±5.8	0.63	2.12±5.6
2	0	0	0.67	1.98±5.1	0.67	1.98±5.1
2	0	1	0.64	2.06±5.7	0.72	2.49±7.1
2	0	2	0.67	2.03±5.6	0.72	2.47±6.9
2	0	3	0.68	2.04±5.5	0.74	2.51±7.1
2	0	4	0.67	2.03±5.6	0.72	2.46±7.0
2	0	5	0.66	2.06±5.6	0.72	2.48±7.0
2	1	0	0.67	1.99±5.1	0.67	1.99±5.1
2	1	1	0.64	2.04±5.7	0.67	2.05±5.5
2	1	2	0.67	2.03±5.6	0.66	1.98±5.2
2	1	3	0.68	2.03±5.4	0.66	2.03±5.3
2	1	4	0.68	2.01±5.5	0.67	2.00±5.3
2	1	5	0.66	2.06±5.6	0.65	2.07±5.6

Table C.41
SVR with RBF kernel with reduced data t_2^* relative error.

			7 day		14 day	
# of Diffs	**Spl.**	**Sam.**	**Med. Rel. Error**	**Mean Rel. Error**	**Med. Rel. Error**	**Mean Rel. Error**
0	0	1	0.85	3.81±10.2	0.82	3.68±9.8
0	0	2	0.83	3.59±9.4	0.81	3.63±10.5
0	0	3	0.78	3.27±8.3	0.79	3.15±7.8
0	0	4	0.80	3.43±8.9	0.83	3.47±9.0
0	0	5	0.78	3.23±8.5	0.79	3.28±8.6
0	1	0	0.78	3.32±8.7	0.80	3.47±8.9
0	1	1	0.83	3.92±10.8	0.84	3.84±10.4
0	1	2	0.83	3.57±9.3	0.83	3.71±10.6
0	1	3	0.79	3.30±8.3	0.80	3.17±7.7
0	1	4	0.82	3.48±9.0	0.83	3.53±9.1
0	1	5	0.82	3.44±9.1	0.79	3.32±8.7
1	0	0	0.75	3.23±8.4	0.75	3.23±8.4
1	0	1	0.88	3.72±9.8	0.84	3.59±9.6
1	0	2	0.83	3.74±9.9	0.81	3.79±10.5
1	0	3	0.86	3.42±8.3	0.87	3.37±8.1
1	0	4	0.81	3.45±8.9	0.81	3.47±9.1
1	0	5	0.78	3.31±8.6	0.79	3.31±8.6
1	1	0	0.78	3.37±8.8	0.78	3.37±8.8
1	1	1	0.86	3.84±10.2	0.83	3.70±9.9
1	1	2	0.83	3.71±9.7	0.83	3.79±10.5
1	1	3	0.86	3.52±8.6	0.88	3.44±8.1
1	1	4	0.82	3.46±9.0	0.82	3.49±9.1
1	1	5	0.80	3.45±9.2	0.79	3.33±8.7
2	0	0	0.81	3.34±8.8	0.81	3.34±8.8
2	0	1	0.87	3.82±10.2	0.87	3.82±10.2
2	0	2	0.82	3.62±9.5	0.82	3.62±9.5
2	0	3	0.82	3.24±8.0	0.82	3.24±8.0
2	0	4	0.85	3.52±8.9	0.85	3.52±8.9
2	0	5	0.85	3.44±8.8	0.85	3.44±8.8
2	1	0	0.80	3.37±8.9	0.80	3.37±8.9
2	1	1	0.87	3.85±10.2	0.88	3.83±10.2
2	1	2	0.84	3.60±9.2	0.82	3.63±9.5
2	1	3	0.82	3.30±8.0	0.84	3.28±8.0
2	1	4	0.84	3.49±8.9	0.85	3.52±8.9
2	1	5	0.82	3.42±8.7	0.86	3.45±8.8

Appendix D

Piecewise Aggregate Approximation Fit

The figures in this sections display the behaviors which are seen for the evaluations being conducted across the folds. Fig. D.1 is the length of survival distributions for each fold used in the 10-fold cross validation and Fig. D.2 is the histogram view.

D.1 Length of Survival Distributions

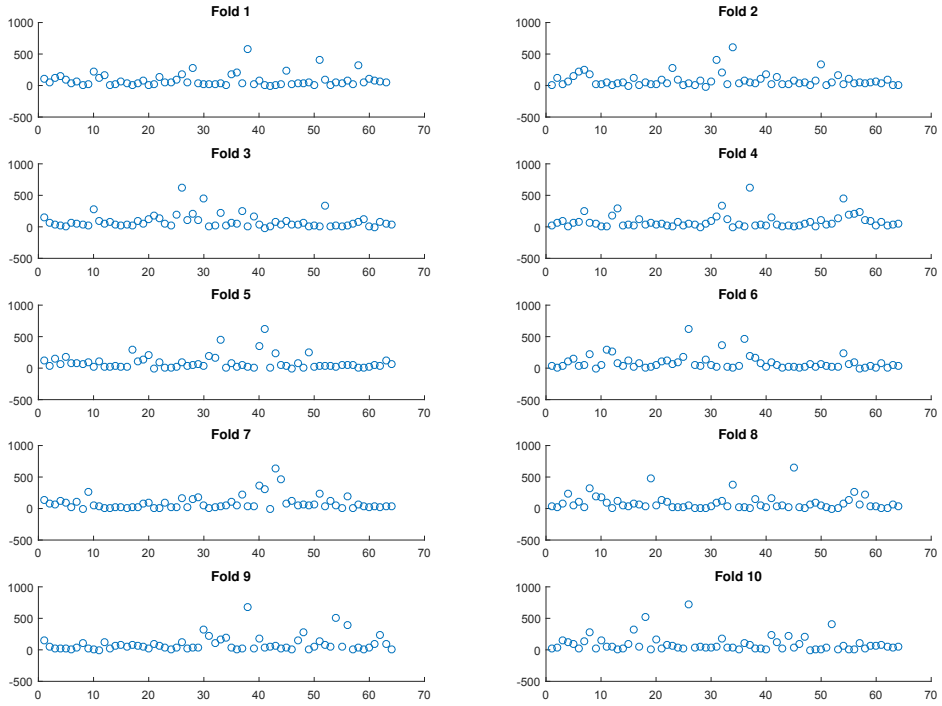


Figure D.1: LOS distributions per fold, all exhibit the same behaviors overall, a product of the cross validation split used.

D.2 PAA Segmentation

In the first phase of segmentation, the genetic algorithm utilized a fitness function which incorporated the evaluation of the model. This is elaborated on in Fig. D.3.

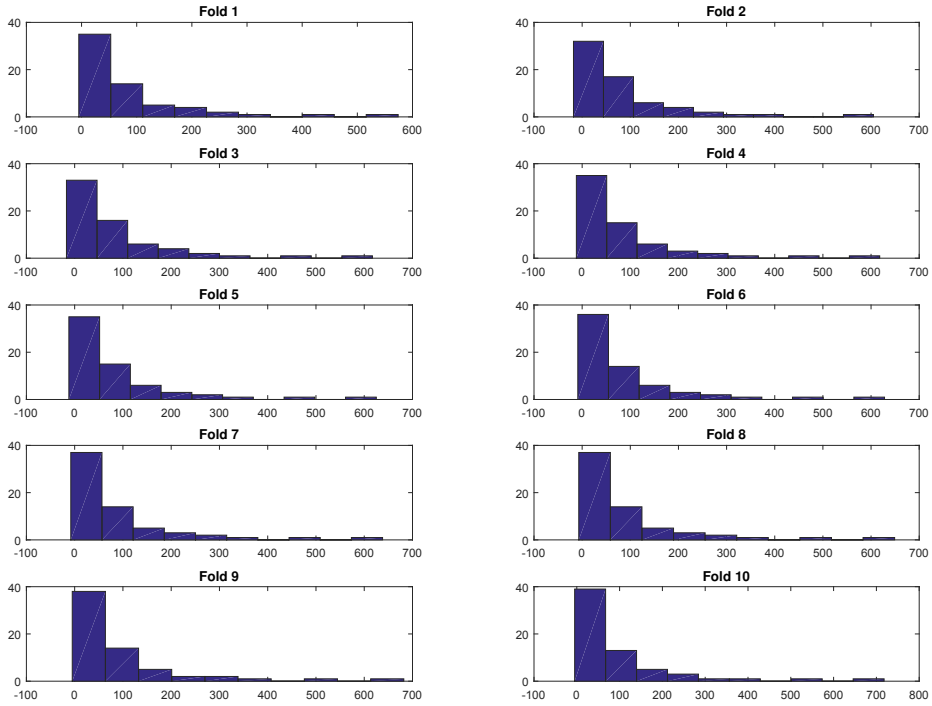


Figure D.2: LOS histograms for censored data set distributions per fold, all exhibit the same behaviors overall, a product of the cross validation split used.

This was not used in the evaluations due to the computational intensity being infeasible. During the evaluation the complete evaluation is needed for each genetic algorithm fitness calculation, with support vector regression this becomes even more intensive with the addition of nested cross validation for parameter selection

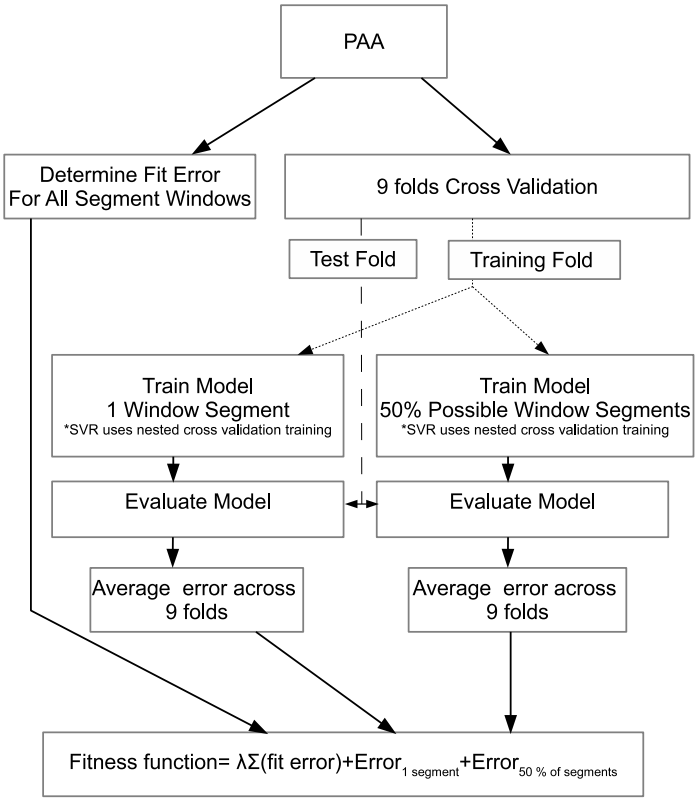


Figure D.3: Fitness function for genetic algorithm initial version.

Appendix E

Copyright Permissions

In this appendix is the correspondences and documentation pertaining to permissions to use work that is published in some variation elsewhere. All of the work has been modified from the form published in these venues.

E.1 Taxonomy of Representation Figure

The correspondence for permission to use the adapted taxonomy of representation figure, Figure 2.4.

Michigan Tech Jennifer Winikus <jawiniku@mtu.edu>

Re: Copyright Permissions

1 message

Jessica Lin <jessica@gmu.edu> Thu, Nov 10, 2016 at 4:19 PM
To: Jennifer Winikus <jawiniku@mtu.edu>

Hi Jennifer,

Congratulations! That's fine regarding the figure.

Regards,
Jessica

Dr. Jessica Lin
Associate Professor
Department of Computer Science
George Mason University
4400 University Drive, MS 4A5
Fairfax, VA 22030
(703) 993-4693
(703) 993-1710 Fax
http://www.cs.gmu.edu/~jessica

"Prediction is very difficult, especially about the future."
-- Niels Bohr

From: Jennifer Winikus <jawiniku@mtu.edu>
Date: Thursday, November 10, 2016 at 4:26 PM
To: Jessica Lin <jessica@cs.gmu.edu>
Subject: Copyright Permissions

Hi Dr. Lin,

I just successfully defended my PhD, and in my Dissertation I have included a modified version of Figure 1 from the paper "A Symbolic Representation of Time Series, with Implications for Streaming Algorithms" by Lin et. al. (doi 10.1145/882082.882086) published in 2003 that you did with Dr. Keogh. I would like permissions to include the modified figure in my dissertation with the citation to the paper. Thank you.

Jennifer Winikus
Teaching Assistant Professor
Department of Computer Science and Engineering
University at Buffalo, The State University of New York
Davis 351 | 716-645-4757 | jwinikus@buffalo.edu
Computer Engineering- PhD Candidate- Michigan Technological University
jawiniku@mtu.edu

E.2 2016 IEEE International Conference on Electro Information Technology

The work presented in Chapter 3 was published in part in the conference proceedings for 2016 IEEE International Conference on Electro Information Technology. Here is the copyright policy granting use in the dissertation.

E.3 Advances in Science, Technology and Engineering Systems Journal

The correspondence and copyright information for the content published in the Advances in Science, Technology and Engineering Systems Journal related to the work in Chapter 4.

Michigan Tech Jennifer Winikus <jawiniku@mtu.edu>

Re: Copyright Permissions

1 message

ASTES Journal <editor@astesj.com> Sun, Oct 16, 2016 at 2:20 PM
To: Jennifer Winikus <jawiniku@mtu.edu>
Cc: Wang Xiu Ying <wangxiuying104@gmail.com>

Although its out of our policy but you are using it for your dissertation, so you can reuse it. I have send you the modified copyright transfer form. I hope you will be satisfied with it while submitting. I have marked the not applicable condition with red texted line.

If you did not feel satisfied with this copyright form, then just leave the copyright transfer form and submit the blank pdf form during camera ready submission of paper.

CC: Prof. Wang Xiu Ying

Best Regards,
Prof. Passerini Kazmerski
Editor-in-Chief
Advances in Science, Technology and Engineering Systems Journal (ASTESJ)
editor@astesj.com
http://www.astesj.com

On 2016-10-17 00:00, Jennifer Winikus wrote:

> Greetings,
>
> I have been accepted to publish with ASTES my paper, "Representation of Clinical Information in Outpatient Oncology for Prognosis Using Regression". I would like to request copyright permission to reuse the content as part of my PhD dissertation at Michigan Technological University.
>
> Thank you.
>
> Jennifer Winikus
> Teaching Assistant Professor
> Department of Computer Science and Engineering
> University at Buffalo, The State University of New York
> Davis 351 | 716-645-4757 | jwinikus@buffalo.edu
> Computer Engineering- PhD Candidate- Michigan Technological University
> jawiniku@mtu.edu

copyright form.docx
13K

Advances in Science, Technology and Engineering Systems Journal

(ISSN: 2415-6698)

Copyright Form

Title of the paper: Representation of Clinical Information in Outpatient Oncology for Prognosis Using Regression

Corresponding author's information (please print):

Name: Jennifer Winikus

Address: 351 Davis Hall
Buffalo, NY 14260

It is herein agreed that:

The copyright to the contribution identified above is transferred to ASTESJ (http://www.astesj.com). This copyright transfer covers the exclusive right to reproduce and distribute the contribution, including reprints, translations, photographic reproductions, microform, electronic form (offline, online), or any other reproductions of similar nature.

~~The Author may not publish his/her contribution anywhere else without the prior written permission of the publisher unless it has been changed substantially. The Author warrants that his/her contribution is original, except for such excerpts from copyrighted works as may be included with the permission of the copyright holder and author thereof, that it contains no libelous statements, and does not infringe on any copyright, trademark, patent, statutory right, or propriety right of others. The Author signs for and accepts responsibility for releasing this material on behalf of any and all coauthors.~~

In return for these rights:

ASTESJ agrees to publish the identified contribution at his own cost and expense. The undersigned hereby gives permission to ASTESJ to publish the above contribution to the above Journal.

Author's Signature: [redacted]

Date: 10/17/2016