UNIVERSITI PUTRA MALAYSIA

REPRODUCING KERNEL HILBERT SPACE METHOD FOR COX PROPORTIONAL HAZARD MODEL

NUR'AZAH BINTI ABDUL MANAF

FS 2016 7

REPRODUCING KERNEL HILBERT SPACE METHOD FOR COX PROPORTIONAL HAZARD MODEL

By

NUR'AZAH BINTI ABDUL MANAF

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfillment of the Requirements for the Degree of Doctor of Philosophy

May 2016

Abstract of this thesis presented to the Senate of Universiti Putra Malaysia in Fulfillment of the requirement for the degree of Doctor of Philosophy

REPRODUCING KERNEL HILBERT SPACE METHOD FOR COX PROPORTIONAL HAZARD MODEL

By

NUR'AZAH BINTI ABDUL MANAF

May 2016

Chair: Associate Professor Mohd. Rizam Bin Abu Bakar, PhD
Faculty: Science

Numerous researchers are enthusiastic about statistical modeling to estimate the survival for patients. Usually, the information obtained from the survival data in biomedical sciences includes the age, race, health conditions, disease free time and the survival times of patients. Apart from developing the survival data models, estimations on the hazard functions are being done to estimate the chance of survival or the time from diagnosis to failure or death of the patients.

It is expected from the result of analysis that using the proposed reproducing kernel method to develop survival models will be helpful in predicting the relapse time or death of patients and will contribute to intense understanding on the connection between reproducing kernels and survival data, which on exploitation will lead to more applications especially in solving related problems in statistics of several areas.

Reproducing kernel Hilbert spaces (RKHS) has been used in the statistics literature for many years. This research explores the mathematical aspects and properties of reproducing kernel Hilbert space. The purpose of this research is to review the basic facts and the importance of RKHS that contribute to the kernel method and its application in statistics by analyzing the effect of kernel method on survival data.

We propose a new reproducing kernel Hilbert space (RKHS) and prove that the kernel obtained satisfy the properties of RKHS. The task is to extend the Cox proportional hazard model by using the new reproducing kernel obtained and apply the kernel method to randomly selected survival data sets. The new kernel we construct will be used in the score function $f(x)$ of the representer theorem for the hazard model. As for the methodology, we obtain the partial differentials of the risk or loss function to fit the hazard model. We find optimal values of parameters of the score function $f(x)$ by using the Newton-Raphson method, which requires setting up the related function to be minimized. Then, we apply the kernel method to the survival data. Finally, we propose an algorithm of minimization of the loss function in the general Cox model. This algorithm is used to determine the vector a_i that enables us to find the optimal parameters of $f(x)$ which is simplified as

$f(x)=\sum_{i=1}^{n} a_i K(x,x_i)$. The survival of patients is estimated through the observation of the exponential values, exp(f(x)) the of model. The $f(x)$ values will affect the risk or failures of the patients. Simulations with different number of covariates will be performed using the proposed kernel $K(\mathbf{A}x, \mathbf{B}y) = \langle \mathbf{A}x, \mathbf{B}y \rangle$. The simulations are done to investigate the effects of different number of covariates on the prediction of overall survival of patients.

We have constructed a new reproducing kernel RKHS and obtained partial differentials of the loss function. The kernel method is expected to be efficient for problems involving data with a large number of covariates. The findings of this research will encourage future exploration of the use of kernel method in the prediction of survival times or failures in many areas such as science, engineering and economics.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

KAEDAH INTI PENJANAAN SEMULA RUANG HILBERT UNTUK MODEL BAHAYA BERKADARAN COX

Oleh

NUR'AZAH BINTI ABDUL MANAF

Mei 2016

Pengerusi : Profesor Madya Mohd. Rizam Bin Abu Bakar, PhD
Fakulti: Sains

Ramai penyelidik telah menunjukkan minat dalam pemodelan statistik untuk mandirian bagi pesakit. Maklumat yang diperoleh daripada data mandirian dalam bidang sains bioperubatan termasuk keadaan kesihatan pesakit seperti tahap tumor, masa bebas penyakit dan masa hidup. Selain membina model data mandirian, anggaran mengenai fungsi bahaya kini dilakukan untuk menganggarkan peluang untuk terus hidup atau masa dari diagnosis kegagalan atau kematian pesakit kanser.

Adalah dijangka dari hasil analisis bahawa penggunaan kaedah inti penjanaan semula untuk membina model mandirian akan membantu dalam ramalan masa berulang atau kematian pesakit dan akan menyumbang kepada pemahaman mendalam terhadap hubungan antara inti penjanaan semula dan data mandirian, yang mana dengan eksploitasinya akan mendorong kepada lebih banyak aplikasi terutama dalam menyelesaikan masalah berkaitan statistik bagi beberapa bidang.

Inti penjanaan semula ruang Hilbert (RKHS) telah digunakan dalam kesusasteraan statistik selama beberapa tahun. Penyelidikan ini meneroka aspek matematik dan sifat-sifat RKHS. Kajian ini mengkaji semula fakta-fakta asas dan kepentingan RKHS yang menyumbang kepada kaedah kernel dan aplikasinya dalam statistik dengan menganalisis kesan kaedah kernel terhadap data mandirian.

Kami mencadangkan inti penjanaan semula ruang Hilbert yang baharu dan membuktikan bahawa inti yang diperolehi memuaskan sifat RKHS. Tugas utama adalah memperluaskan model bahaya berkadaran Cox dengan menggunakan inti penjanaan semula yang baharu diperolehi dan menggunakan kaedah kernel ke atas data mandirian yang dipilih secara rawak. Kernel baharu yang kami bina akan digunakan di dalam fungsi skor $f(x)$ dalam teorem perwakilan bagi model bahaya. Pengkaedahannya adalah memperolehi pembezaan separa fungsi risiko atau fungsi kerugian untuk disesuaikan dengan model bahaya yang digunakan. Kami mencari nilai-nilai optimum bagi parameter fungsi skor $f(x)$ dengan menggunakan kaedah Newton-Raphson yang memerlukan pembentukan fungsi berkaitan untuk diminimumkan dan Seterusnya, kami menggunakan kaedah kernel untuk data

mandirian. Akhir sekali, kami mencadangkan algoritma untuk meminimumkan fungsi bahaya model umum Cox. Algoritma ini digunakan untuk menentukan vector a_i yang membolehkan kita untuk mencari parameter optimum bagi $f(x)$ yang dipermudahkan sebagai $f(x) = \sum_{i=1}^{n} a_i K(x, x_i)$. Mandirian pesakit dianggar melalui pemerhatian nilai-nilai eksponen bagi model. Nilai-nilai $f(x)$ akan memberi kesan kepada risiko atau kegagalan pesakit. Simulasi dengan beberapa bilangan kovariat akan dilaksanakan dengan inti kernel $K(\mathbf{A}x, \mathbf{B}y) = \langle \mathbf{A}x, \mathbf{B}y \rangle$ yang dicadang. Simulasi dilakukan untuk menyiasat kesan the bilangan kovariat yang berbeza terhadap ramalan mandirian pesakit.

Kami telah menjana satu inti inti penjanaan semula ruang Hilbert (RKHS) dan memperolehi pembezaan separa fungsi kerugian. Kaedah kernel ini berkesan untuk masalah yang melibatkan data dengan bilangan kovariat yang besar. Hasil kajian ini akan menggalakkan penerokaan masa depan penggunaan kaedah kernel dalam ramalan masa hidup atau kegagalan dalam banyak bidang seperti sains, kejuruteraan dan ekonomi.

ACKNOWLEDGEMENT

First and foremost, I would like to thank the Almighty for the good health, peace of mind, strength, and wisdom He bestowed upon me in order to perform my research and to complete my thesis writing.

First and foremost, I would like to thank the Almighty for the good health, peace of mind, strength, and wisdom He bestowed upon me in order to perform my research and to complete my thesis writing.

I would like to express my deep gratitude to the Supervisory Committee Members: Associate Professor Dr. Mohd. Rizam Abu Bakar, Associate Professor Dr. Gafurjan Ibragimov and Professor Dr. Habshah Midi for their guidance, encouragement and insightful comments.

I am highly indebted to Associate Professor Dr. Gafurjan Ibragimov who has been always there to listen and give advice to continue and complete my doctorial study. I am deeply grateful to him for the long discussions that helped me sort out the details of my work especially on functions and equations, and at the same time helped me to enrich my ideas. Through him I also learned the proper way to present my research findings and results. His patience and continuous support helped me to overcome many critical situations and finish this thesis. I am also thankful to him for carefully reading and commenting on countless revisions of this thesis.

My appreciation to Dr. Bader Al Jawadi whom as a graduate student had willingly assisted and guided me in the R programming language. Many thanks to Dr. Zahriddin Muminov for providing necessary information regarding my research to enhance my thesis contents. I am grateful to many friends who have helped me stay sane through these difficult years. Their moral support and care helped me overcome setbacks and stay focused on my doctorial study.

Most importantly, my success would not have been possible without the love and patience of my family. My very special thanks to my beloved late mother Datin Alawiah Omar who passed away in the midst of my study, in 2012, whom till the end of her days reminded me to be determined to accomplish my Ph.D. degree. I would like to express my heart-felt gratitude to my father, Dato' Abdul Manaf Abdul Hamid and my extended family members who had aided and encouraged me throughout this endeavour. My deep appreciation goes to my husband, Abdul Rahman Thambi Ibrahim who has been a constant source of concern, support and pillar of strength all these years. I am grateful to my children Hana, Amar, Zhafri, Hani and Dina for their understanding and tolerance during my difficult times in my struggle of completing this doctorial study.

Finally, I appreciate the financial support from Ministry of Higher Education and Universiti Teknologi Malaysia that funded four years of my study.

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy.

The members of the Supervisory Committee were as follows:

Mohd. Rizam Bin Abu Bakar, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Chairman)

Gafurjan Ibragimov, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Member)

Habshah Binti Midi, PhD
Professor
Faculty of Science
Universiti Putra Malaysia
(Member)

BUJANG KIM HUAT, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

Declaration by graduate student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any other institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature: ________________________ Date: ___________________

Name and Matric No.: Nur'azah Binti Abdul Manaf (GS19184)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Many researchers have shown interest in statistical modeling of survival for patients. Information obtained from the survival data in biomedical sciences includes the patients' health conditions such as the stage of the tumor, disease free time and the survival time. Besides developing the survival data models, estimations on the hazard functions are being done to estimate the chance of survival or the time from diagnosis to failure or death of the cancer patients.

Several studies have been done on the survival data by using the reproducing kernels as mentioned by Aronszajn (1950), Hille (1972), Burbea (1976), Wahba (1998), Berlinet et al. (2003). It is expected from the result of analysis that using the proposed reproducing kernel method to develop survival models will be helpful in predicting the relapse time or death of patients and will contribute to intense understanding on the connection between reproducing kernels and survival data, which on exploitation will lead to more applications especially in solving related problems in statistics of several areas.

1.1. Background

The aim of this research is to use a reproducing kernel function in Hilbert space over the field of real numbers to estimate the hazard function for survival models.

The mathematical aspects and properties of reproducing kernel in Hilbert space (RKHS) is explored in this research to understand basic facts and the importance of RKHS that contribute to the kernel method and its application in statistics are being reviewed. It is known that kernel methods provide a framework for solving several profound issues in the theories of machine learning. A combination of kernel techniques, machine learning theory, and optimization algorithms contribute to the development of kernel-based learning methods. Some reproducing kernels used in survival analysis will be introduced to show the importance of reproducing kernel method in the area of science and statistics.The mathematical concepts of Newton-Raphson method and the numerical methods for function optimization in statistics will be discussed. The function $f(x)$ of the representer theorem that involves the reproducing kernels is obtained by generating the mathematical process behind this method. The process of finding the solution to the regularised least-squares problem via a system of linear equations is illustrated to explain the procedures to find the values of parameters involved in the kernel method.

Several approaches to survival analysis using Cox's proportional hazards regression modelling in the attempts to model the time until an event of interest has been reported in literature of survival analysis. Smith et al. (2000) stated that the robust nature of the Cox proportional hazard model allows close approximation of the result for the correct parametric model when comparing the hospitalization experiences of two or more cohorts. In the research study the authors found a model which suggested an increase in risk among the exposed patients during hospitalization. The reproducing kernels are usually used to extend the Cox regression model and in the application to the survival data. Kernel Cox regression models were introduced by Li (2003) for relating expression profiles to censored survival data and applied to three types of cancer data sets. The simple natural inner product kernel $K(x_i, y_j) = \langle x_i, x_j \rangle$ was used in the research.

The important facts related to reproducing kernels need to be explained before extending the discussion on the research.

1.2 Kernels

A kernel is a symmetric continuous function, $K:[a,b]\times[a,b]\rightarrow\mathbb{R}$, so that $K(x,s)=K(s,x)$ (Scholkopf, 2002). In order to further understand about reproducing kernels and its properties we need to know the basic concepts that contribute to kernel learning factors. Kolmogorov (1941) started to study the methods for representing kernels in linear spaces for a countable output domain. The method for representing kernels in linear spaces for general cases was developed by Aronszajn (1950).

1.2.1 Positive Definite Matrix

Definition 1.1 (Gentle, 2007)
A $p\times p$ symmetric matrix A is said to be *positive definite* if, for all vectors $x\in\mathbb{R}^p$, the quadratic form $x^T Ax$ is positive, that is

$$x^T Ax > 0,\ x \neq 0.$$

Suppose that $A=\left[a_{ij}\right]=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{bmatrix}$ is a positive definite matrix. Then,

matrix A has the following properties (Gentle, 2007).

1. The $r\times r$ submatrix $A_r=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1r} \\ a_{21} & a_{22} & \cdots & a_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ a_{r1} & a_{r2} & \cdots & a_{rr} \end{bmatrix}$, where $1\leq r\leq p$ is also positive

   definite.

2. All the p eigenvalues of A, $\lambda_1, \lambda_2, \ldots, \lambda_p$ are positive. If all the eigenvalues of a matrix are positive, the matrix is also a positive definite matrix.

3. A unique decomposition of A, $A = LL^T$ exists, where L is a lower triangular matrix, that is

$$L = \left[L_{ij} \right] = \begin{bmatrix} l_{11} & 0 & \cdots & 0 \\ l_{21} & l_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{p1} & l_{p2} & \cdots & l_{pp} \end{bmatrix}.$$

   The equation $A = LL^T$ gives the *Cholesky Decomposition* of *A*.

4. A unique decomposition of *A*, $A = SS$ exists, where $S = \sqrt{A}$ is the matrix square root of *A*.

5. A unique decomposition of A, $A = VDV^T$ exists, where

$$D = diag(\lambda_1, \lambda_2, \ldots, \lambda_p) = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{bmatrix}$$

   is the diagonal matrix composed of the eigenvalues of *A*, and *V* is the orthogonal matrix.

6. By properties 2 and 5, as $A = VDV^T$, $|V| = 1$, and $|D| = \prod_{i=1}^{p} \lambda_i > 0$, we have

$$|A| = \left|VDV^T\right| = |V||D|\left|V^T\right| = |V|^2 |D| = |D| > 0.$$

7. By property 6, because $|A| > 0$, then *A* is non-singular which means that the inverse of *A*, A^{-1} exists such that

$$AA^{-1} = A^{-1}A = 1.$$

   Subsequently, we can deduce that

$$A^{-1} = \left(VDV^T\right)^{-1} = VD^{-1}V^T,$$

   since $V^{-1} = V^T$.

8. The inverse of a matrix *A*, A^{-1} is also positive definite.

9. For $x \in \mathbb{R}^p$,

$$\min_{1 \le i \le p} \lambda_i \le \frac{x^T A x}{x^T x} \le \max_{1 \le i \le p} \lambda_i .$$

It is known (Stewart, 1976) that positive definite matrices are related closely to positive-definite symmetric bilinear forms, and to inner products of vector spaces. Saitoh (1988) showed the connection between the positivity or *positive matrix* defined by Aronszajn (1950) and the positive semi-definiteness of all finite set kernel matrices.

1.2.2 Positive Definite Kernels

It is obvious that positive definite kernel is a generalization of a positive matrix in operator theory. It provides a framework to the construction of basic Hilbert spaces.

Definition 1.2

Let $L(H_i, H_j)$ be the bounded operators from H_i to H_j and $\left\{H_n\right\}_{n \in \mathbb{R}}$ be a sequence of real Hilbert spaces. Then, a map A on $Z \times Z$ where $A(i, j)$ lies in $L(H_i, H_j)$ is called a **positive definite kernel** if for all $k > 0$, the following positive definiteness condition hold (Hille, 1972):

$$\sum_{-k \le i,\, j \le k} \left\langle A(i, j) h_j , h_i \right\rangle \ge 0$$

where $h_i \in H_i$, $h_j \in H_j$.

1.2.3 Gram Matrix

Gram matrix is named after a Danish Mathematician Jorgen Pederson Gram (Hazewinkle, 2001). The Gram Matrix of a set of vectors $x_1, x_2, \ldots, x_m$ in an inner product space is the Hermitian matrix of inner products, whose entries are given by $K_{ij} = \left\langle x_i, x_j \right\rangle = K\left(x_i, x_j\right)$ and is defined as following:

Definition 1.3

Given a function $K : X^2 \to \mathbb{R}$ or $K : X^2 \to \mathbb{C}$ and patterns $x_1, x_2, \ldots, x_m \in X$. The $m \times m$ matrix K with elements $K_{ij} = \left\langle x_i, x_j \right\rangle = K\left(x_i, x_j\right)$ is called the Gram matrix or kernel matrix with respect to $x_1, x_2, \ldots, x_m$.

Given a real matrix A, the matrix $A^T A$ is a Gram matrix (of the columns of A), while the matrix AA^T is the Gram matrix of the rows of A. Kernel functions are often represented by Gram matrices (Lanckriet, 2004).

1.2.4 Cauchy-Schwarz Inequality for Kernels

The Cauchy–Schwarz inequality is one of the most important inequalities in mathematics and is very useful inequality encountered in different settings, such as analysis, linear algebra, probability and statistical theory. It is an inequality which has a number of generalizations to solve problems.

The Cauchy–Schwarz inequality for any vectors x and x' of an inner product space is normally written stated as (Scholkopf, 2002):

$$\left|\langle x, x' \rangle\right| \leq \|x\| \|x'\|.$$

Equivalently, in terms of inner product, the inequality can be written as

$$\langle x, x' \rangle^2 \leq \langle x, x \rangle \langle x', x' \rangle,$$

which is known as the Cauchy-Schwarz inequality for kernels.

1.3 Scope

This research focused mainly on the reproducing kernel Hilbert space (RKHS) and the kernel method that will be applied to selected survival data. The properties of RKHS are being explored thoroughly in order to construct a new RKHS that will be used in the generalized Cox hazard model of the kernel method. The properties are verified upon constructing the new RKHS before it is used in the survival model. The important goal is to show that RKHS plays an important role as a tool in kernel method and its application to survival analysis.

1.4 Problem Statements

The use of the reproducing kernel Hilbert space (RKHS) is of interest because RKHS provides a base for adaptable function estimation and statistical modelling with direct, indirect and scattered data distributions. According to Li and Luan (2003), models based on RKHS are foundation for penalized likelihood estimation and regularization methods and can handle wide variety of data distributions and problems. In statistics, RKHS can be used to estimate the survival of patients' data with many covariates. Solutions of optimization problems in RKHS are highly important tools in many fields of mathematical investigations for engineers, computer scientists and statisticians.

The reproducing kernels, the properties of RKHS and the kernel method are not commonly utilized by researchers in Malaysia. Several reproducing kernels in Hilbert space had been used extensively by international researchers but locally the researches on RKHS focus mostly on theoretical aspects and rarely used to solve problems in areas of applied mathematics, statistics and engineering. It is interesting being able to construct new functions based on the properties of existing functions and compare the result of solutions with the other commonly used functions. Once the new function is constructed, we should be able to apply the function in mathematical or statistical modelling to solve problems in several areas of research. The important task to obtain solutions for randomly selected data set is to obtain the appropriate algorithm and equations for a model.

1.5 Research Objectives

The research is conducted to achieve the following objectives:

- *To construct a new reproducing kernel.*

  A new reproducing kernel Hilbert space is constructed and proven that the kernel obtained satisfies the properties of RKHS. The new kernel will be constructed using two diagonal matrices. We must show that the newly constructed kernel satisfy the positive definiteness and symmetry properties.

- *To find optimal values of parameters of the score function $f(x)$.*

  The mathematical procedures to find the optimal values of parameters for survival data will be explained. In order to find solution using the Newton-Raphson method, we have to set up the related function to be minimized and obtain the partial differentials to fit the hazard model used in this research.

- *To apply the kernel method to the survival data.*

  The Cox regression model is extended by using the new reproducing kernel and the kernel method is applied to the survival data. The new kernel we constructed will be used in the score function $f(x)$ of the representer theorem of the hazard model.

- *To propose an algorithm to minimize the loss function in the general Cox model*.

  The algorithm is used to determine the vector a_i that enables us to find the optimal parameters of $f(x)$ which is simplified as $f(x) = \sum_{i=1}^{n} a_i K(x, x_i)$. The value of $f(x)$ will determine the value of $\exp(f(x_i))$ which is the factor of the Cox proportional hazard model.

1.6 Outlines of Thesis

The research work is started with the exploration of the RKHS and its properties. Upon understanding all aspects of reproducing kernel Hilbert space, a new RKHS is constructed and the properties that classify the kernel as RKHS is shown. Once the new kernel is constructed, an initial exploratory data analysis is performed by using the negative partial log likelihood function as the loss function. This process is done by minimizing the loss function to find the optimal parameter values of survival data for the kernel method. The Newton-Raphson method is used to solve the optimization problem. Survival data of HIV positive patients from a public hospital is used in the application of the new modified kernel method. The exponential values of the kernel model will be observed to estimate the survival of patients.

This thesis is organized as follows:
In Chapter 1, we introduce the research by stating the background. The important facts related to reproducing kernels are explained. We also highlighted the problem statements and objectives of the research.

Literature review is discussed in Chapter 2. In this chapter, we will review the previous work done by researchers that are related to the reproducing kernel Hilbert space (RKHS), the Cox proportional hazard function, the utilization of kernel method in statistics and the application of kernel method in medical sciences.

In Chapter 3, we provide the mathematical backgrounds and the fundamental tools in this research such as Hilbert space, reproducing kernel Hilbert space and its properties, survival analysis and Cox proportional hazard model. Several definitions, theorems and examples are included.

In chapter 4, we will focus on the methodology of the research that will lead to the fulfillment of the objectives mentioned in Chapter 1. We begin the chapter with the construction of the new reproducing kernel. Then, we show that the new kernel satisfy the properties of the reproducing kernel Hilbert space: positive definite and symmetry. We then show the use of the new kernel in the link function $f(x)=\sum_{i=1}^{n} a_i K(x, x_i)$ and extend to the Cox model. Subsequently, we will explain the procedures to obtain the partial differentials to enable us to find the optimal values of the parameters of the score functions. At the end of the chapter, we discuss the values of $\exp(f(x))$ to estimate the survival time of the patients.

In Chapter 5, we will show and discuss the results of the construction of the new kernel, the simplified equations, and obtain partial differentials of functions to find the optimal values of parameters for the Cox model. We compare the solutions we get from the application of the new kernel with the other kernels: linear kernel, quadratic kernel and Gaussian Radial Basis Function Gaussian RBF) kernel.

In Chapter 6, we include the summary and general conclusions of the research. In this chapter, we suggest the construction of several new reproducing kernels in Hilbert space and gave recommendations for future research. Lastly, we recommend the application of kernel methods in area engineering, medical science, economics and finance.

REFERENCES

Abu Bakar, M.R., Salah, K. A. and Ibrahim, N.A. (2009). Bayesian Approach for Joint Longitudinal and Time-to-Event Data with Survival Fraction. *Bulletin of the Malaysian Mathematical Science Society* (2), Vol. 32(1): 75-100.

Alpay, D., Jorgensen, P. and Volok, D. (2014). Relative Reproducing Kernel Hilbert Spaces. *Proceeding of American Mathematical Society*, 142 (11):3889-3895.

Alpay, D., Jorgensen, P., Seager, R. and Volok, D. (2013). On Discrete Analytic Functions: Products, Rational Functions and Reproducing Kernels. Journal of Applied Mathematics and Computing, Vol. 41(1): 393–426.

Alpay, D. and Dym, H.(1993). On a New Class of Structured Reproducing Kernel Spaces, *Journal of Functional Analysis*. 111 (1) :1-28.

Alpay, D., Bolotnikov, V., Dijksma, A. and Snoo, H. (1993). On Some Operator Colligations and Associated Reproducing Kernel Hilbert Spaces, Operator Extensions, Interpolation of Functions and Related Topics. *Oper. Theory Adv. Appl*., Vol. 61: pp. 1–27.

Alpay, D. and Dym, H.(1992). On Reproducing Kernel Spaces, the Schur Algorithm, and Interpolation in a General Class of Domains, Operator theory and Complex Analysis. *Oper. Theory Adv. Appl*., vol. 59: 30 -77.

Applebaum, D. (2004). Lévy Processes and Stochasitic Calculus. Cambridge Studies in Advanced Mathematics. Cambridge Univ. Press, Cambridge, UK.

Arasan, J. and Lunn, M. (2009). Survival Model of a Parallel System with Dependent Failures and Time Varying Covariates. *Journal of Statistical Planning and Inference*, Vol. 139(3): 944-951.

Aronszajn, N. (1950). Theory of Reproducing Kernels. *Transaction of American. Mathematical Soc*iety 68: 337- 404.

Bergman, S. (1950). The Kernel Function and Conformal Mapping. *Mathematical Surveys and Monographs*, 257.

Berlinet, A. and Thomas-Aqnan, C. (2003). Reproducing Kernel Hilbert Spaces in Probability and Statistics. *Kluwer Academic*, U.K.

Bertero, M, Poggio, T.A.,Torre,V. (1988). Ill-posed Problems in Early Vision, *Proceeding IEEE* 76, 869–889.

Blei, D.M., Jordan, M.I. (2006). Variational Inference for Dirichlet Process

Mixtures. *Bayesian Anal*., 1(1):121–143 (electronic).

Breslow, N.E. (1974). Covariance Analysis of Censored Survival Data. *Biometrics*, 30: 89-99.

Burbea, J. (1976). Total Positivity of Certain Reproducing Kernels. *Pacific Journal of Mathematics*, *67(1)*: 101-130.

Caponnetto, A., Micchelli, C.A., Pontil, M. and Ying, Y. (2008). Universal Multi-task Kernels. *Journal of Machine Learning Research*, Volume 9: 1615–1646.

Cao, H., Churpek, M.M., Zeng, D. and Fine, J.P. (2015). Analysis of the Proportional Hazards Model With Sparse Longitudinal Covariates, *Journal of the American Statistical Association*, 110:511:1187-1196, DOI: 10.1080/01621459.2014.957289

Casquilho, M. (2000). Useful Mathematical Tools. *SIAM News: Volume 33(4)*: 7-10.

Cawley, G.C., Talbot, N.LC (2003). Efficient Leave-One-Out Cross-Validation of Kernel Fisher Discriminant Classifiers. *Pattern Recognition*, Volume 36(11): 2585-2592.

Cawley, G.C., Talbot, N.L.C., Janacek, G.J., Peck, M.W. (2004). Bayesian Kernel Learning Methods for Parametric Accelerated Life Survival Analysis. *Deterministic and Statistical Methods in Machine Learning*, 37-55.

Cawley, G.C., Talbot, N.LC., Janacek, G.J. , Peck, M.W. (2006). Sparse Bayesian Kernel Survival Analysis for Modeling the Growth Domain of Microbial Pathogens. IEEE Transactions on Neural Networks, 17(2):471-481.

Cawley, G.C.(2006). Leave-One-Out Cross-Validation Based Model Selection Criteria for Weighted LS-SVMs. *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, pp. 1661 - 1668.

Chakraborty,S., Ghosh, M., and Mallick. B.K. (2005). Bayesian Non-linear Regression for Large p Small n Problems. *Journal of Multivariate Analysis*, 108: 28-40.

Cox, D.R. (1972). Regression Models and Life-Tables (with discussion). *Journal of the Royal Statistical Society – Series B*, 34: 187-220.

Cox, D.R., Oakes, D. (1984). Analysis of Survival Data. *Monographs on Statistics and Applied Probability, Volume 21*. Chapman and Hall.

Cristianini, N., and Shawe-Taylor, J. (2000). An Introduction to Support Vector Machines. Cambridge University Press.

Cucker, C. and Smale, S. (2002). On the Mathematical Foundations of Learning, Bulletin of. American Mathematical Society, 39 (1): 1-49 (electronic). MR 1864085 (2003a:68118).

Efron, B. (1977). The Efficiency of Cox's Likelihood Function for Censored Data. *Journal of the American Statistical Association 72:* 557-565.

Ferguson, T.S. (1973). A Bayesian Analysis of Some Non-parametric Problems. *Ann. Stat*., 1:209–230.

Ferguson, T.S. (1974). Prior Distributions on Spaces of Probability Measures. *Ann. Stat.*, 2:615–629.

Geng, Y., Lu, W. and Zhang, H. (2014). A Model-Free Machine Learning Method for Risk Classification and Survival Probability Prediction. *Stat.* 3(1): 337–350. doi:10.1002/sta4.67.

Gentle, J.E. (2007). Matrix Algebra: Theory, Computations, and Applications in Statistics. Springer Science & Business Media, New York. pp. 277-281.

Ghosal, S. and Roy, A. (2006). Posterior Consistency of Gaussian Process Prior for Non-Parametric Binary Regression. Ann. Statist., 34(5): 2413-2429.

Girosi, F., Jones, M. and Poggio, T. (1990) Networks and the Best Approximation Property. *Biol. Cybern*. 63: 169-176.

Girosi, F., Jones, M. and Poggio, T. (1993). Priors, Stabilizers and Basis Functions: From Regularization to Radial, Tensor and Additive Splines. *Technical Report A.I. Memo No. 1430*, MIT.

Girosi, F., Jones, M. and Poggio, T. (1995). Regularization Theory and Neural Network Architectures. *Neural Computation*, 7:219–269.

Girosi, F.(1998). An Equivalence between Sparse Approximation and Support Vector Machines. Neural Computation. 10(6):1455–1480.

Hansen, V. L. (2006). Functional Analysis: Entering Hilbert Space. World Scientific Publication Co. Inc.

Hazewinkel, M. (2001). Gram matrix. *Encyclopedia of Mathematics*. Springer.

Hedenmalm, H. and Nieminen,P.J. (2014). The Gaussian Free Field and Hadamard's Variational Formula. *Probability Theory and Related Fields*, 159(1): 61–73.

Hille, E. (1972). Introduction to General Theory of Reproducing Kernels. *Rocky Mountain Journal of Mathematics, 2(3)*: 321-367.

Hooshang, H. (2010). *Tribology of Interface Layers*. CRC Press. ISBN 978-0- 8247-5832-5.

Ishwaran H., Kogalur U.B., Glackstone, E.H., and Lauer M.S. (2008). Random Survival Forests. *The Annals of Applied Statistics*, 2:841–860.

James, L.F., Lijoa, A., and Pr¨unster, I. (2005). Conjugacy as a Distinctive Feature of the Dirichlet Process. *Scandinavian Journal of Statistics*, 33:105-120.

Jorgensen, P and Tian, F. (2015). Graph Laplacians and Discrete Reproducing Kernel Hilbert Spaces from Restrictions and Cameron-Martin. arXiv: 1501.04954 v1 [math.FA].

Kalbfleisch, J.D. and Prentice, R.L. (2002). The Statistical Analysis of Failure Time Data. Wiley-Interscience.

Kallianpur, G. (1970). The Role of Reproducing Kernel Hilbert Spaces in the Study of Gaussian Processes. Advances in Probability and Related Topics, 2:49–83.

Kimeldorf, G. and Wahba, G. (1971). Some Results on Tchebycheffian Spline Functions. *Journal of Mathematical Analysis and Applications*, 33:82–95.

Kiani, K. and Arasan, J. (2012). Simulation of Interval Censored Data in Medical and Biological Studies. International Conference Mathematical and Computational Biology 2011 International Journal of Modern Physics: Conference Series,Vol. 9: 112–118.

Klein, J.P. and Moeschberger, M.L.(1997). Survival Analysis: Techniques for Censored and Truncated Data. New York, Springer.

Kulkarni, S. and Harman, G. (2011). An Elementary Introduction to Statistical Learning Theory. Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., Hoboken, NJ.

Lanckriet, G. R. G. et al. (2004). Learning the Kernel Matrix with Semi-Definite Programming. *Journal of Machine Learning Research 5*: 27–72.

Lata, S. and Paulsen, V. (2011) The Feichtinger Conjecture and Reproducing Kernel Hilbert Spaces, *Indiana Univ. Mathematics Journal*, 60 (4): 1303–1317.

Lee E.T, Go O.T. (1997). Survival analysis in public health research. *Annual Review Public Health*, 18:105-134.

Levitan, B.M. (2001). Hilbert Space. Encyclopedia of Mathematics, Springer.

Li, H., Luan Y. (2003). Kernel Cox Regression Models for Linking Gene Expression Profiles to Censored Survival Data. *Pacific Symposium on Bio-Computing 8*: 65 – 76.

Liang et.al. (2007). Understanding the Use of Unlabelled Data in Predictive Modeling. *Statistical Science*, Volume 22, Number 2: 189-205.

Liang et.al.,(2007). Non-parametric Bayesian Kernel Models. Discussion Paper 2007-10, Duke University ISDS, Durham, NC, URL : emwww.stat.duke.edu/research/papers/.

Lin, Y. and Brown, L.D. (2004). Statistical Properties of the Method of Regularization with Periodic Gaussian Reproducing Kernel, Annals of Statistics, 32(4): 1723–1743.

Loève, M. (1948). Fonctions Aleatoires du Second Ordre. Processus Stochastiques et Mouvement Brownien, pp. 299-352.

Long, H., Zhang, X. (2009). Construction and Calculation of Reproducing Kernel Determined by Various Linear Differential Operators. *Applied Mathematics and Computations*, 215(2009):759-766.

Lukic', M.N., Beder, J.H. (2001). Stochasitic Processes with Sample Paths in Reproducing Kernel Hilbert Spaces. *T. Am. Math. Soc*., 353(10):3945– 3969.

MacKay, D.J.C. (1992). Bayesian Interpolation. *Neural Computation 4*, pp. 415 - 447.

MacKay, D.J.C. (1992). A Practical Bayesian Framework for Backprop Networks. *Neural Computation 4*: 448-472.

MacKay, D.J.C. (1992). The Evidence Framework Applied to Classification Networks. *Neural Computation 4*: 720–736.

Mercer, J.(1909). Functions of Positive and Negative Type and their Connection with the Theory of Integral Equations. *Philosophical Transactions Royal Society London*. A209:415 - 446.

Mika, S., RÄatsch, G., Weston, J., SchÄolkopf, B., and MÄuller, K.R. (1999). Fisher Discriminant Analysis with Kernels. *Neural Networks for Signal Processing,*
Volume IX : 41 - 48. IEEE Press, New York.

Montgomery, D.C., Peck, E.A. (1992). Introduction to Linear Regression Analysis. John Wiley & Sons, 2nd edition.

Moore, E. H., Barnard, R.W. (1935). General Analysis. *Memoirs of the American Philosophical Society, I.* American Philosophical Society, Philadelphia, Pennsylvania, pp. 197-209.

Parzen, E. (1961). An approach to time series analysis. Annual Math. Stat., 32: 951-989.

Parzen, E. (1963). Probability Density Functionals and Reproducing Kernel Hilbert Spaces. *Proceedings of the Symposium on Time Series Analysis.* New York, NY, 1963. John Wiley & Sons, pp. 155-169.

Prinja, S., Gupta, N., Verma, R. (2010). Censoring in Clinical Trials: Review of Survival Analysis Techniques. Indian Journal of Community Medicine, 35(2): 217–221.

Quang, M. H., Kang, S.H. and Le, T. M. (2010). Image and Video Colorization Using Vector-valued Reproducing Kernel Hilbert Spaces. *Journal of Mathematical Imaging and Vision,* 37 (1): 49–65.

Rasmussen, C.E., Williams, C.K.I. (2006). Gaussian Processes for Machine Learning. MIT Press, Cambridge, MA.

Reed, M. and Simon, B. (1972). Methods of Modern Mathematical Physics. I: Functional Analysis (v.1). Academic Press.

Scholkopf, B., Smola, A.J. (2002). Learning with Kernels. The MIT Press.

Schölkopf, B., Muandet, K., Fukumizu, K., Harmeling, S. and Jonas Peters, J. (2015).Computing Functions of Random Variables via Reproducing Kernel Hilbert Space Representations. *Statistics and Computing*, 25(4): 755–766. DOI 10.1007/s11222-015-9558-5.

Smale, S. and Zhou, D. (2009). Online learning with Markov Sampling, Anal. Appl.(Singap.) 7, No. 1: 87–113.

Smith,T.C., Gray G.C.,Knoke J.D. (2000). Is Systeic lupus Erythematosus Amyotrophic lateral Sclerosis or Fibromyalgia associated with Persian Gulf War Service? An examination of Department of Defence Hospitalization. *American Journal of Epidermiology*, Vol 151.

Smith, T., Smith, B. (2001). Survival Analysis and the Application of Cox's Proportional Hazards Modeling using SAS. *Proceedings of the Twenty-sixth Annual SAS User's Group International Conference*. SAS Institute Inc., pp.

244-246.

Smola, A.J., Scholkopf, B., Muller, K.R. (1998). The Connection between Regularization Operators and Support Vector Kernels. *Neural Networks,* 11:637–649.

Stewart, J. (1976). Positive Definite Functions and Generalizations, an Historical Survey. *Rocky Mountain Journal of Mathematics*, 6(3).

Taweab, F., Ibrahim, N.A. and Arasan, J. (2015). A Bounded Cumulative Hazard Model with a Change-point According to a Threshold in a Covariate for Right-censored Data. Applied Mathematics & Information Sciences, Vol. 9(1): 69-74.

Tobar, F.A. and Mandic, D.P. (2014). Quaternion Reproducing Kernel Hilbert Spaces: Existence and Uniqueness Conditions. IEEE Transactions on Information Theory, Vol. 60, No. 9: 5736-5749.

Tu et.al., (2006). Lévy Adaptive Regression Kernels. Discussion Paper 2006-08, Duke University ISDS, Durham, NC, 2006. URL http://www.stat.duke.edu/research/papers/.

Vapnik, V.N. (1998). The Nature of Statistical Learning Theory. Springer, New York, 2nd edition.

Vuletić, M. (2013). The Gaussian Free Field and Strict Plane Partitions. Alain Goupil and Gilles Schaeffer.25th International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC 2013), Paris, France. *Discrete Mathematics and Theoretical Computer Science* (*DMTCS Proceedings*), AS, pp.1041-1052.

Wahba, G. (1990). Splines Models of Observational Data. Series in Applied Mathematics, Vol. 59. SIAM, Philadelphia.

Wahba, G. (1998). Technical Report, UW Madison.

Wahba, G. (1999). Support Vector Machines, Reproducing Kernel Hilbert Spaces, and Randomized GACV. In B. Sch¨olkopf, C.J.C. Burges, and A.J. Smola, editor, *Advances in Kernel Methods - Support Vector Learning*. The MIT Press, Cambridge, MA, pp. 69–88.

Wahba, G. (2003). An Introduction to Reproducing Kernel Hilbert Spaces and Why are They so Useful. Proceedings of the 13th IFAC Symposium on System Identification (SYSID 2003).

Wang, Z. and Wang, C.Y. (2010), Buckley-James Boosting for Survival Analysis with High-Dimensional Biomarker Data. *Statistical Applications in Genetics and Molecular Biology*, Vol. 9 (1), Article 24.

Williams, C.K.I. (1998). Prediction with Gaussian processes: From Linear Regression to Linear Prediction and Beyond. *Learning and Inference in Graphical Models*. Kluwer pp. 599–621.

Wolpert et.al. (2003). A Nonparametric Bayesian Approach to Inverse Problems (with discussion). *Bayesian Statistics 7*: 403–418, Oxford Univ. Press. ISBN 0-19-852615-6.

Wu, C. (1980). Linear Algebra and its Applications, Volume 32, pages 1-234.

Xing et.al. (2004). Bayesian Haplotype Inference via the Dirichlet Process. Machine Learning, Proceedings of the 21st International Conference (ICML 2004), Banff, Canada, New York, NY. ACM Press. URL http://www. aicml.cs.ualberta.ca/_banff04/icml/pages/accepted.htm.

Xing et.al. (2006). Bayesian Multi-Population Haplotype Inference via a Hierarchical Dirichlet Process Mixture. *Proceedings of the 23rd International Conference (ICML 2006)*, Pittsburgh, PA, New York, NY. ACM Press. http://www.icml2006.org/ icml2006/technical/accepted.html.

Zhang, H., Xu,Y. and Zhang, Q. (2012) Refinement of Operator-valued Reproducing Kernels, Journal of Machine Learning Research, Vol. 13: 91-136.