UNIVERSITI PUTRA MALAYSIA

ROBUST DIAGNOSTICS AND PARAMETER ESTIMATION METHODS IN LINEAR AND NONLINEAR REGRESSION BASED ON NU SUPPORT VECTOR REGRESSION FOR HIGH DIMENSIONAL DATA

AL-DULAIMI ABDULLAH MOHAMMED RASHID

IPM 2022 5

ROBUST DIAGNOSTICS AND PARAMETER ESTIMATION METHODS IN LINEAR AND NONLINEAR REGRESSION BASED ON *NU* SUPPORT VECTOR REGRESSION FOR HIGH DIMENSIONAL DATA

By

AL-DULAIMI ABDULLAH MOHAMMED RASHID

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

June 2022

DEDICATION

I would like to dedicate this dissertation work to

My dear mother, who taught me the meaning of patience, her prayers and beautiful words have always helped me during this period of study.

My dear father, who taught me the meaning of perseverance in this life, has always been a source of support and encouragement to me.

My dear sisters, your love and beautiful words have always supported me.

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

ROBUST DIAGNOSTICS AND PARAMETER ESTIMATION METHODS IN LINEAR AND NONLINEAR REGRESSION BASED ON *NU* SUPPORT VECTOR REGRESSION FOR HIGH DIMENSIONAL DATA

By

AL-DULAIMI ABDULLAH MOHAMMED RASHID

June 2022

Chairman : Professor Habshah binti Midi, PhD
Institute : Mathematical Research

Support Vector Regression (SVR) has become increasingly popular in the detection of outliers and classification problems in high dimensional data (HDD), because it can handle nonlinear, rank deficient and high dimensional problems by employing the kernel trick to transform nonlinear relationship in the input space into a linear form in a high dimensional feature space.

The standard SVR and the μ-ε-SVR are introduced to detect outliers in HDD. Nonetheless, they are computationally expensive and not very successful in detecting outliers. As a solution to this problems, the fixed parameters support vector regression (FP-ε-SVR) was put forward. The FP-ε-SVR using ε-SVR is also not very successful in identifying outliers. A *nu*-SVR is developed to overcome these shortcomings. The results signify that the proposed *nu*-SVR method is very successful in identifying outliers under a variety of situations, and with less computational running time.

The statistically inspired modification of the partial least squares (SIMPLS) is the widely used method to handle partial least squares problems in high dimensional data. However, the SIMPLS is no longer efficient when outliers are present in the data. The robust iteratively reweighted SIMPLS (RWSIMPLS) technique, which is an enhancement of the SIMPLS algorithm, is put forward to remedy this problem. Nevertheless, with regard to parameter estimations and outlier diagnostics, the RWSIMPLS is still inefficient. It also suffers from long computational times. Hence, a new robust RWSIMPLS (SVR-RWSIMPLS) algorithm that incorporates a new weight function constructed from *nu*-SVR, is established. The numerical results clearly indicate the SVR-RWSIMPLS algorithm is more efficient, more robust and has less computational running times than the RWSIMPLS. The proposed SVR-RWSIMPLS diagnostic plot is also very successful in classifying observations into correct groups.

The least absolute shrinkage and selection operator (LASSO) is the shrinkage procedure based on penalized function which is the commonly used method in performing parameter estimations and variables selection, simultaneously. However, the LASSO method is easily affected by outliers since it is a special case of the penalized least squares regression with L_1 penalty function, where L_1 is the regularization penalty parameter that limits the regression coefficients' size, such that some coefficients can become zero and be eliminated. Many penalization methods have been proposed to remedy this problem that includes the WLAD-LASSO. However, the shortcoming of the WLAD-LASSO is that its efficiency tends to decrease as the number of good leverage points (outlying observations in X-space where they follow the pattern of the majority of the data) increases. Moreover, it can only handle low dimensional data because it is based on the robust mahalanobis distance - minimum volume ellipsoid (RMD-MVE) weight whereby MVE can only be computed for low dimensional data. Thus, a new weighted WLAD-LASSO method (SVR-WLL) is developed to simultaneously estimate the parameters and variables selection of regression model. The results of simulation study and real data sets show that the SVR-WLL is superior compared to the existing methods discussed in this thesis.

The Principal component analysis (PCA) is the most commonly used approach for analysing high dimensional data in order to achieve dimension reduction. However, outliers have an adverse effect on the PCA, hence reduce the accuracy of the prediction model. To date, no research has been done to incorporate SVR technique in the algorithm of PCA in order to obtain accurate prediction model with high accuracy. To close the gap in the literature, a new hybrid PCA with the *nu*-SVR technique (SVR-PCA) is established. The results show that the SVR-PCA is more efficient than the PCA technique.

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

DIAGNOSTIK TEGUH DAN KAEDAH ANGGARAN PARAMETER DALAM REGRESI LINEAR DAN BUKAN LINEAR BERDASARKAN VEKTOR SOKONGAN REGRESI *NU* UNTUK DATA BERDIMENSI TINGGI

Oleh

AL-DULAIMI ABDULLAH MOHAMMED RASHID

Jun 2022

Pengerusi : Profesor Habshah binti Midi, PhD
Institut : Penyelidikan Matematik

Regresi Vektor Sokongan (SVR) telah menjadi semakin popular dalam pengesanan titik terpencilr dan masalah klasifikasi dalam data dimensi tinggi (HDD), kerana ia boleh menangani masalah bukan linear, kekurangan pangkat dan dimensi tinggi dengan menggunakan helah kernel untuk mengubah hubungan tak linear dalam input ruang ke dalam bentuk linear dalam ruang ciri dimensi tinggi.

Kaedah SVR biasa dan μ-ε-SVR diperkenalkan untuk mengesan titik terpencil dalam HDD. Walau bagaimanapun, pengiraannya rumit dan tidak begitu berjaya dalam mengesan titik terpencil. Sebagai penyelesaian kepada masalah ini, parameter tetap penyokong regresi vektor (FP-ε-SVR) dikemukakan. Kaedah FP-ε-SVR menggunakan ε-SVR juga tidak begitu berjaya dalam mengenal pasti titik terpencil. Kaedah Nu-SVR dibangunkan untuk mengatasi kelemahan ini. Hasilnya menunjukkan bahawa kaedah nu-SVR yang dicadangkan sangat berjaya dalam mengenal pasti titik terpencil di bawah pelbagai situasi, dan mengambil masa pengiraan yang pantas.

Pengubahsuaian statistik kuasa dua terkecil separa (SIMPLS) ialah kaedah yang digunakan secara meluas untuk menangani masalah kuasa dua separa terkecil dalam data dimensi tinggi. Walau bagaimanapun, SIMPLS tidak lagi cekap apabila terdapat titik terpencil dalam data. Teknik SIMPLS (RWSIMPLS) pemberat semula yang teguh, yang merupakan peningkatan algoritma SIMPLS, dikemukakan untuk menyelesaikan masalah ini. Namun begitu, RWSIMPLS masih tidak cekap dari segi anggaran parameter dan diagnostik terpencil,. Ia juga mengalami masa pengiraan yang panjang. Oleh itu, algoritma RWSIMPLS (SVR-RWSIMPLS) teguh baharu yang menggabungkan fungsi berat baharu yang dibina daripada *nu*-SVR, diwujudkan. Keputusan berangka jelas megambil masa pengiraan yang kurang daripada RWSIMPLS. Plot diagnostik SVR-

RWSIMPLS yang dicadangkan juga sangat berjaya dalam mengklasifikasikan pemerhatian kepada kumpulan yang betul.

Pengendali pengecutan dan pemilihan mutlak terkecil (LASSO) ialah prosedur pengecutan berdasarkan fungsi berhukum yang merupakan kaedah yang biasa digunakan dalam melaksanakan anggaran parameter dan pemilihan pembolehubah, secara serentak. Walau bagaimanapun, kaedah LASSO mudah dipengaruhi oleh titik terpencil kerana ia adalah kes khas regresi kuasa dua terkecil yang dikenakan penalti dengan fungsi penalti L_1, di mana L_1 ialah parameter penalti regularisasi yang menghadkan saiz pekali regresi, supaya beberapa pekali boleh menjadi sifar dan dihapuskan. Banyak kaedah penalti telah dicadangkan untuk membetulkan masalah ini termasuk WLAD-LASSO. Walau bagaimanapun, kelemahan WLAD-LASSO ialah kecekapannya cenderung menurun apabila bilangan titik tuasan tinggi yang baik (pemerhatian terpencil dalam ruang-*X* yang mengikut corak majoriti data) meningkat. Selain itu, ia hanya boleh mengendalikan data dimensi rendah kerana ia berdasarkan pemberat bagi jarak mahalanobis yang teguh – isipadu minimum ellipsoid (RMD-MVE) di mana MVE hanya boleh dikira untuk data dimensi rendah. Oleh itu, kaedah WLAD-LASSO berwajaran baharu (SVR-WLL) dibangunkan untuk menganggarkan pemilihan parameter dan pembolehubah model regresi secara serentak. Hasil kajian simulasi dan set data sebenar menunjukkan bahawa SVR-WLL adalah lebih unggul berbanding kaedah sedia ada yang dibincangkan dalam tesis ini.

Analisis komponen utama (PCA) ialah pendekatan yang paling biasa digunakan untuk menganalisis data dimensi tinggi untuk mencapai pengurangan dimensi. Walau bagaimanapun, titik terpencil mempunyai kesan buruk pada PCA, oleh itu mengurangkan ketepatan model ramalan. Sehingga kini, tiada kajian telah dilakukan untuk menggabungkan teknik SVR dalam algoritma PCA bagi mendapatkan model ramalan yang tepat dengan ketepatan yang tinggi. Untuk merapatkan jurang dalam kesusasteraan, PCA hibrid baharu dengan teknik nu-SVR (SVR-PCA) dibangunkan. Keputusan menunjukkan bahawa SVR-PCA adalah lebih cekap daripada teknik PCA.
Vektor Sokongan Regresi (SVR) telah menjadi semakin popular dalam pengesanan data terpencil dan masalah klasifikasi dalam data berdimensi tinggi (HDD), kerana ia boleh menangani masalah tak linear, kekurangan pangkat dan dimensi tinggi oleh penggunaan helah kernel untuk mengubah hubungan tak linear di dalam ruang input kepada bentuk linear dalam ruang ciri berdimensi tinggi.

Piawaian SVR dan μ-ε-SVR telah diperkenalkan untuk mengesan titik terpencil dalam HDD. Namun, ianya adalah mahal dari segi pengiraan dan tidak begitu berjaya dalam mengesan data terpencil. Sebagai penyelesaian kepada masalah ini, parameter tetap Vektor Sokongan Regresi (FP-ε-SVR) telah dikemukakan. FP-ε-SVR menggunakan ε-SVR juga tidak begitu berjaya dalam mengenal pasti data terpencil. *nu*-SVR dibangunkan untuk mengatasi kelemahan tersebut. Keputusan menunjukkan bahawa cadangan kaedah *nu*-SVR sangat berjaya dalam mengenal pasti titik terpencil di bawah pelbagai situasi, dan dengan masa pengiraan yang kurang.

Pengubahsuai yang diilhamkan secara statistik bagi kuasa dua terkecil separa (SIMPLS) ialah kaedah yang digunakan secara meluas untuk menangani masalah kuasa dua terkecil separa dalam data berdimensi tinggi. Walaubagaimanapun, SIMPLS tidak lagi cekap dengan kehadiran titik terpencil dalam data. Teknik perulangan pemberat teguh SIMPLS (RWSIMPLS) yang merupakan peningkatan algoritma SIMPLS, dikemukakan untuk menyelesaikan masalah ini. Namun, berkenaan dengan anggaran parameter dan diagnostik outlier, RWSIMPLS masih tidak cekap. Ia juga mengalami masa pengiraan yang panjang. Maka, algoritma teguh baru RWSIMPLS (SVR-RWSIMPLS) yang menggabungkan fungsi pemberat baharu yang dibina daripada *nu*-SVR, diwujudkan. Keputusan berangka dengan jelas menunjukkan SVR-RWSIMPLS algoritma adalah lebih cekap, lebih teguh dan mempunyai masa pengiraan yang kurang daripada RWSIMPLS. Plot diagnostik SVR-RWSIMPLS yang dicadangkan juga sangat berjaya dalam mengklasifikasikan pemerhatian ke dalam kumpulan yang betul.

Pengecutanc mutlak terkecil dan pemilihan pengendali (LASSO) ialah prosedur pengecutan berdasarkan fungsi penalti yang merupakan kaedah yang biasa digunakan dalam melaksanakan anggaran parameter dan pemilihan pembolehubah, secara serentak. Namun, kaedah LASSO mudah dipengaruhi oleh data terpencil kerana ia adalah kes khas penalti regresi kuasa dua terkecil dengan fungsi penalti L_1. Banyak kaedah penalti telah dicadangkan untuk membetulkan masalah ini termasuk WLAD-LASSO. Walaubagaimanapun, kelemahan WLAD-LASSO ialah kecekapannya cenderung menurun apabila bilangan titik tuasan yang baik (pemerhatian terpencil dalam ruang-*X* di mana ia mengikut corak majoriti data) meningkat. Selain itu, ia hanya boleh mengendalikan data berdimensi rendah, ($p < n$) kerana ia berdasarkan berat RMD-MVE di mana MVE hanya boleh dikira untuk data berdimensi rendah. Oleh itu, kaedah WLAD-LASSO berwajaran baharu (SVR-WLL) dibangunkan secara serentak menganggar pemilihan parameter dan pembolehubah model regresi. Hasil kajian simulasi dan set data sebenar menunjukkan bahawa SVR-WLL adalah lebih unggul berbanding kaedah sedia ada yang dibincangkan dalam tesis ini.

Analisis komponen utama (PCA) ialah pendekatan yang paling biasa digunakan untuk menganalisis data berdimensi tinggi untuk mencapai pengurangan dimensi. Walau bagaimanapun, titik terpencil mempunyai kesan buruk terhadap PCA, oleh itu mengurangkan ketepatan model ramalan. Sehingga kini, tiada kajian telah dilakukan untuk menggabungkan teknik SVR dalam algoritma PCA bagi mendapatkan model ramalan yang tepat dengan ketepatan yang tinggi. Untuk merapatkan jurang dalam literatur, PCA hibrid baharu dengan teknik *nu*-SVR (SVR-PCA) dibangunkan. Keputusan menunjukkan bahawa SVR-PCA adalah lebih cekap daripada teknik PCA.

ACKNOWLEDGEMENTS

Firstly, I would like to thank the Almighty Allah for the ability, patience, knowledge, and strength bestowed on me to conduct this study.

Sincere thanks, appreciation, and respect to the supervisors of Prof. Dr. Habsha Midi. I greatly benefited from her guidance and support. I am very honored to have had the opportunity to complete my degree under her supervision.

I would also like to thank my co-supervisors, Associate Prof. Dr. Jayanthi A/P Arasan and Dr. Mohd Shafie Bin Mustafa for all their supporting and guidance provided.

My special thanks to my external supervisor Dr. Waleed Dhhan for his guidance, encouragement, and support. Also, I would also like to express my thanks and gratitude to Associate Prof. Dr. Hassan S. Uraibi for his continuous support.

I would like to thank my family for their support, prayers and encouragement.

This thesis was submitted to the Senate of the 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:

Habshah binti Midi, PhD
Professor
Faculty of Science
Universiti Putra Malaysia
(Chairman)

Jayanthi a/p Arasan, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Member)

Mohd Shafie bin Mustafa, PhD
Senior Lecturer
Faculty of Science
Universiti Putra Malaysia
(Member)

Waleed Dhhan, PhD
Associate Professor
Scientific Research Centre
Nawroz University, Iraq
(Member)

ZALILAH MOHD SHARIFF, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 10 November 2022

Declaration by Members of Supervisory Committee

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision;
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature:	
Name of Chairman of Supervisory Committee:	Professor Dr. Habshah binti Midi
Signature:	
Name of Member of Supervisory Committee:	Associate Professor Dr. Jayanthi a/p Arasan
Signature:	
Name of Member of Supervisory Committee:	Dr. Mohd Shafie bin Mustafa
Signature:	
Name of Member of Supervisory Committee:	Associate Professor Dr. Waleed Dhhan

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

A-LASSO	Adaptive LASSO
BIF	Bounded Influence Function
BLUE	Best Linear Unbiased Estimator
BP	Breakdown Point
DRGP	Diagnostic Robust Generalized Potential
DV	Dependent Variable
FP-SVR	Fixed Parameter Support Vector Regression
GM	Generalized M estimators
HDD	High Dimensional Data
HLP	High Leverage Point
IF	Influence Function
IID	Independent Identically Distributed
IV	Independent Variables
KKT	Karush Kuhn Tucker
LAD	Least Absolute Deviation
LASSO	Least Absolute Shrinkage Selection Operator
LAV	Least Absolute Values
LMS	Least Median of Squares
LSM	Least Squares Method
LTS	Least Trimmed Squares
MCD	Minimum Covariance Determinant
MD	Mahalanobis Distance

MSVR	Modified Support Vector Regression
MVE	Minimum Volume Ellipsoid
NNR	Neural Network Regression
OLS	Ordinary Least Squares
PCA	Principal Component Analysis
PLSR	Partial Least Squares Regression
RBF	Radial Basis Function
RMD	Robust Mahalanobis Distance
RWSIMPLS	Robust Iteratively Reweighted SIMPLS
SIMPLS	Statistically Inspired Modification of the PLS
SLT	Statistical Learning Theory
SSVR	Standard Support Vector Regression
SVC	Support Vector Classification
SVM	Support Vector Machine
SVR	Support Vector Regression
WLAD	Weighted Least Absolute Deviation

CHAPTER 1

INTRODUCTION

1.1 Background of the Study

Regression analysis resembles a statistical procedure for determining the functional relationship between two or more variables in order to predict a dependent variable (DV) (output) from one or more independent variables (IV) (input) (Kutner et al., 2005). Given the explanatory variables, regression analysis calculates the response variable's conditional expectation. That means when the independent variables are fixed, it calculates the average value of the dependent variable. This can be accomplished by employing the appropriate approach for the data set under investigation. One of the most prevalent estimation strategies in regression analysis is the ordinary least squares method (OLS). The OLS is among the common estimation technique in the linear regression, due to its attractive properties and ease of calculation. Moreover, provided that random errors are independent and identically normally distributed (IID), then the OLS estimator is the best linear unbiased estimator (BLUE). In the sense that within all possible linear unbiased estimators, the OLS estimator possesses the minimum variance. However, in the vast majority of real-world applications, the assumptions underlying the linear regression model such as the error terms are independent and normally distributed, are not met. Additionally, the OLS estimator is not resilient in the presence of outliers, which are common in real-world scenarios. To put it another way, the OLS estimator has a low breakdown point of 1/n (Maronna et al., 2006), in which n represents the sample size. This implies that even if one point is abnormal, it can drastically affect the least squares estimate in the incorrect direction (refer to Rousseeow and Leroy, 1987; Kamruzzaman and Imon, 2002; Maronna et al., 2006).

In the presence of one or more outlying observations, the normal distribution of the error terms are easily offended. According to Belsley et al. (1980), outliers refer to observations that have the greatest effect on the calculated values of various estimates, either alone or in combination with multiple other points. Additionally, Hawkins (1980) described an outlier as one which differs very significantly than the others that it raises concerns that it was caused by a variety of factors. "An outlier is an observation that, because it is unusual and/or unjustified, deviates decisively from the overall behaviour of experimental data in regards to the criterion studied," as explained by Muoz-Garcia et al. (1990). As per Barnett and Lewis (1994), outliers are points that are significantly different from the bulk of observations included in a data collection. In general, outliers in regression issues may be divided into numerous categories. Outliers, or also called as vertical outliers, are those outlying data in the Y-direction. On the other hand, high leverage points (HLPs) are observations that are outlying in the X-direction, while residual outliers are observations that have large residuals. According to Midi et al. (2021), HLPs can be classified into good and bad leverage points. The good leverage points are located on the regression line, and they possess minor effects on the regression estimators, with the potential to improve the precision of an estimate. On the contrary,

the bad leverage points are located distant from the regression line and possess a large impact on regression estimators.

With regards to the non-linearity relationship and outliers between variables, additional major difficulties that impact the projected model include high-dimensional and sparse (p is much larger than the sample size, n). Additionally, Yatchew (2003) pointed that the potential approximation error is larger ten times for each additional explanatory variable to the regression model because of the difficulty to meet the assumptions of the parametric regression model. Nowadays, there are numerous techniques that deal with these issues independently. Even so, there are few studies in the literature that take into account the existence of outliers, high-dimensional and non-linearity issues (less than full or full rank) all at the same time. Consequently, finding options that provide the essential flexibility to deal with these difficulties, for instance, nonparametric approaches, particularly learning machines, has become a crucial requirement. Therefore in this thesis, the focus is mainly on developing several statistical methods based on support vector regression on high dimensional data. Hence, it is important to first introduce the concept of support vector regression, followed by some other important concepts.

1.2 Support Vector Machine for Regression

The conventional regression assumptions, for instance, the assumption of linear relationship among variables with the knowledge of the data's underlying probability distribution, are difficult to achieve in most real-world situations (Ukil, 2007). Due to this problem, other alternative methods such as the nonparametric machine learning should be adopted.

Support Vector Machine (SVM) is a promising and somewhat new approach for learning to separate functions in classification problems (SVC) or conducting function estimation in regression issues (SVR). Moreover, because of its high performance and capability to transform non-linear relationships among variables into the linear forms using the kernel notion (kernel function), SVM applications have been rising recently. Additionally, Cortes and Vapnik (1995) developed the support vector machine using statistical learning theory (SLT) for distribution-free data learning. They described SVM as a collection of related supervised learning approaches that can be used to solve regression and classification issues. By virtue of its exceptional performance in a diversified learning situation, it has sparked great attention in the machine learning field, both theoretically and empirically. It is good at solving problems in image analysis (Guo et al., 2008), bioinformatics (Ben-Hur et al., 2008), financial prediction and marketing database (Ukil, 2007), bankruptcy prediction (Härdle et al., 2011), artificial intelligence (Frohlich and Zell, 2005), and text categorization (Kuo and Yajima, 2010). Other factors that contribute to the SVM's widespread use include its theoretical assurances regarding its performance and less sensitivity to local minima.

The SVM was first used to solve classification problems (Cortes and Vapnik, 1995). However, the formulation was immediately improved to solve regression issues (Vapnik, 1995; Vapnik et al., 1996). Moreover, the SVM is distinguished by its capacity to generate a comprehensive, sparse, and unique solution (Ceperic et al., 2014). Vapnik's ε-insensitive SVR is a popular support vector regression formulation. It generates a predictive model that is based on only a portion of the training data (sparse model) and disregards any aspects that fall under the threshold ε. The sparse regression model is a simplified model, in general. When compared to the complexity ratio, it can achieve excellent accuracy. On the other hand, the SVR model incorporates a regularisation term (weight) in its training formulation that helps to reduce the model's complexity. Several advantages of the sparse regression model have been identified by several researchers, which may be summarized in the next three points (Tipping, 2001; De Figueiredo, 2003; Roth, 2004; De Brabanter et al., 2010; Guo et al., 2010):

Tendency to avoid the issue of over-fitting: a model with less complexity is less likely to over-fit the data.

Lower computational expenses during active use: the SV machine model's estimate time is significantly related to the support vectors (the support vectors number), and as this number reduces, the execution speed increases.

Capability to generalize: generalization, as well as over-fitting are two notions that are closely associated in that if the likelihood of over-fitting is reduced, the model's ability to generalize is strengthened.

Nevertheless, the SVR's ability to generate a sparse model by itself is insufficient to guarantee that the model will generalize effectively. As an example, when the parameter ε value is too small, then, the generated model will be dependent on the majority of the training points, which leads to the non-sparse resulting solution (Guo et al., 2010).

1.2.1 The Basic Idea

Let consider the training data $\{(x_1, y_1), \ldots, (x_n y_n)\} \subset X \times R$, in which X denotes the input variables space (e.g., $X=R^d$). The purpose of the ε-tube SVR is to search *f(x)* such that it has at most, ε deviation from the acquired outputs y_i. Moreover, at the same time it should be as flat as possible (Smola and Schölkopf, 2004). It is worth mentioning that the training errors are not much taken into consideration as long as they are smaller than the threshold value ε, but any deviation more than this will not be tolerated. Let us first consider the situation of a linear function *f*, as follows:

$$f(x) = \langle w, x \rangle + b, \tag{1.1}$$

in which $\langle .,. \rangle$ describes the dot product in X, $b \in R$ and $w \in X$ denotes the offset and slope of the regression function. Flatness in the function (1.1) means that one attempts a small value by minimizing the Euclidean norm $\|w\|^2$.

This problem may be expressed as a convex optimization problem (Smola and Schölkopf, 2004).

$$\begin{aligned} &minimize && \frac{1}{2}\|w\|^2 \\ &subject\ to && \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon\,. \end{cases} \end{aligned} \tag{1.2}$$

In (1.2), the implied assumption was that the convex optimization problem is feasible. The slack variables are introduced to deal with the issue of the optimization problem's infeasible constraints (1.2). According to Vapnik (1995), this process leads to the formulation:

$$\begin{aligned} &minimize && \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \\ &subject\ to && \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, i = 1,2 \ldots., n\ , \end{cases} \end{aligned} \tag{1.3}$$

in which ξ_i^* and ξ_i resemble the slack variables that provide the lower and the upper errors, and *C* is realized as the tradeoff between model complexity and the number of deviations higher than ε which can be tolerated.

This is equivalence of minimizing the ε-insensitive loss function as given in (1.4), in which it describes the best robustness characteristics among different common loss functions, for instance, Huber's, Gaussian, and Laplacian (Schölkopf and Smola, 2002; Rojo-Álvarez et al., 2003; Colliez et al., 2006).

$$L_\varepsilon\,(y_i) = \begin{cases} 0 & \text{if } |y_i - f(x)| \leq \varepsilon \\ |y_i - f(x)| - \varepsilon & otherwise. \end{cases} \tag{1.4}$$

The hypothetical aspect is visually depicted in Figure 2.1. It can be observed from this figure, that only the points that are falled beyond the ε-tube (the shaded region) are considered as support vectors, and their deviations are penalized in a linear way. As per Lee and Mangasarian (2001), the optimization problem (1.3) can easily be solved in its dual formulation, provided that the dimension of the parameter *w* is much bigger compared to the number of samples. Furthermore, the dual formulation is the key to extend the SVR method to nonlinear functions.

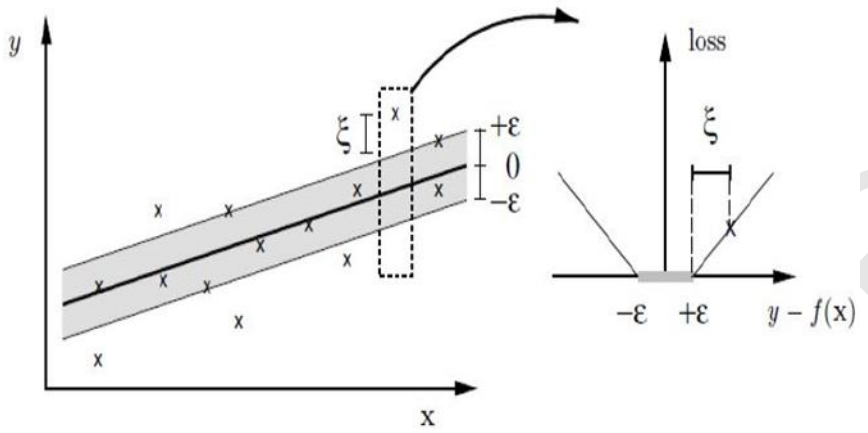

Figure 1.1 : A linear SVM soft margin loss setting (Schölkopf and Smola 2002)

1.2.2 Dual Problem and Quadratic Programs

The main idea is to construct the Lagrange function (L) from the objective function and its accompanying constraints that possesses a saddle point at the solution in terms of the dual and primal variables (for more details refer Vanderbei, 1999). This can be accomplished by introducing dual set of variables that can be shown as follows:

$$\begin{aligned} L = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i - \xi_i^*) - C\sum_{i=1}^{n}(\eta_i\xi_i + \eta_i^*\xi_i^*) \\ -\sum_{i=1}^{n}\alpha_i(\varepsilon + \xi_i - y_i + \langle w, x_i\rangle + b) \\ -\sum_{i=1}^{n}\alpha_i^*(\varepsilon + \xi_i^* + y_i - \langle w, x_i\rangle - b, \end{aligned} \tag{1.5}$$

where $\alpha_i, \alpha_i^*, \eta_i, \eta_i^*$ resemble the Lagrange multipliers that have to meet positivity constraints, such that

$$\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \geq 0.$$

To achieve the optimality, the partial derivatives of L to the primal variables (w, b, ξ_i, ξ_i^*) are selected and be equated to zero (Smola and Schölkopf, 2004).

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{n}(\alpha_i^* - \alpha_i) = 0 \tag{1.6}$$

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)x_i = 0 \tag{1.7}$$

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \eta_i = 0 \tag{1.8}$$

$$\frac{\partial L}{\partial \xi_i^*} = C - \alpha_i^* - \eta_i^* = 0. \tag{1.9}$$

The following dual optimization problem is obtained by substituting (1.6), (1.7), (1.8), and (1.9) into (1.5) (Smola and Schölkopf, 2004):

$$\begin{aligned} maximize \quad & \begin{cases} -\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\langle x_i, x_j\rangle \\ -\varepsilon\sum_{i,j=1}^{n}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{n} y_i(\alpha_i - \alpha_i^*) \end{cases} \\ subject\ to \quad & \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0 \ \ and \ \ \alpha_i, \alpha_i^* \epsilon\, [0, C]. \end{aligned} \tag{1.10}$$

Conditions (1.8) and (1.9) are used to remove the dual variables η_i and η_i^* in deriving (1.10). The weight vector *w* can be obtained by rewriting equation (1.7) as

$$w = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)x_i. \tag{1.11}$$

Hence, the regression function is represented as

$$f(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\langle x_i, x\rangle + b. \tag{1.12}$$

It important to highlight that the parameter *b* can be calculated by exploiting the Karush–Kuhn–Tucker criteria (KKT) (Keerthi et al., 2001; Ceperic et al., 2014). Moreover, the conditions of KKT to nonlinear programming generalize the Lagrange multiplier approach to permit inequality constraints as well as equality requirements (Boyd and Vandenberghe, 2004).

1.2.3 Generalized SVR Algorithm for Nonlinear Case

Only the linear regression situation has been explored in the preceding subsections. The following stage is to build the SVM algorithm nonlinear. This could be achieved by

simply applying the function $\Phi : X \rightarrow F$ on the training patterns x_i. The function Φ is employed to transform the input space X into some feature space F by applying the conventional SV regression procedure (Smola and Schölkopf, 2004; Ceperic et al., 2014). Regrettably, for both high-dimensionality and polynomial (high order) features, this strategy can quickly become computationally expensive (Vapnik, 1995). Clearly, this strategy is not appropriate in all circumstances, and one may look for other methods that are computationally less expensive.

As previously stated, the SVM method is only relies on the dot products among patterns x_i. Thus, Boser et al. (1992) have deduced that knowing $k(x_i, x) = \langle \Phi(x_i), \Phi(x) \rangle$ instead of Φ explicitly is sufficient, allowing us to recast the optimization problem of SVM (1.10) as the following:

$$maximize \begin{cases} -\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i - \alpha_i^*)\left(\alpha_j - \alpha_j^*\right)k(x_i, x_j) \\ -\varepsilon\sum_{i,j=1}^{n}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{n} y_i(\alpha_i - \alpha_i^*) \end{cases}$$

$$subject\ to \quad \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0 \ \ and \ \ \alpha_i, \alpha_i^* \epsilon\, [0, C].$$

Similarly, the equations (1.11) and (1.12) can rewritten as

$$w = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\Phi(x_i),$$

and

$$f(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)k(x_i, x) + b.$$

The distinction between the two scenarios is that in the nonlinear case, unlike the linear example, *w* is not explicitly provided. It is worth noting that, in the nonlinear situation, the optimization issue corresponds to the discovering the flattest function in feature space instead of input space.

1.2.4 The Steps of SVR Algorithm

The different phases of the regression process are visually illustrated in this section. Figure 2.2 shows how the map function Φ is utilized to map the input pattern x_i into the feature space (Schölkopf and Smola, 2002). Then, within the training patterns that were

previously mapped by the map function, compute the dot products. This is equivalent to analysing the kernel functions $k(x_i, x) = \langle \Phi(x_i), \Phi(x) \rangle$. Afterwards the weights $v_i = \alpha_i - \alpha_i^*$ are used to insert the dot products. Lastly, the regression's final prediction output is obtained by adding the parameter b. It is also worth noting that the procedure which has been previously explained is quite comparable to Neural Network Regression (NNR), with the difference that in the case of SV, the weights in the input layer are the training patterns' subset.

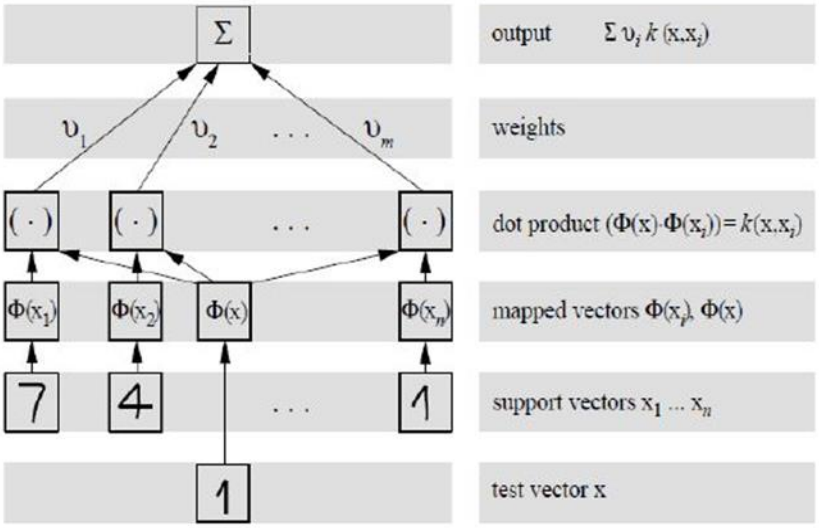

Figure 1.2 : A regression machine's architecture generated by the SV algorithm (Schölkopf and Smola 2002)

1.2.5 The standard SVM Regression for Outlier Detection

With regard to Karush-Kuhn-Tucker (KKT) circumstances, the conventional SVR approach for outlier detection (Jordaan and Smits, 2004) takes advantage of the Lagrange multipliers produced by solving optimization problem. This explains why the output between the dual variables and constraints must dematerialize at the moment of solution.

$$\alpha_i(\varepsilon + \xi_i - y_i + \Phi(x_i) + b) = 0$$
$$\alpha_i^*(\varepsilon + \xi_i^* + y_i - \Phi(x_i) - b) = 0$$
$$\xi_i(C - \alpha_i) = 0$$
$$\xi_i^*(C - \alpha_i^*) = 0\,.$$

If the slack variable is zero for any point, while the upper-bound Lagrange multiplier α_i or α_i^* is not present, then, the data point is not suspected as an outlier. On the other hand, the data points that possess upper bounds Lagrange multipliers α_i and α_i^* may be termed as elected outliers. Because different data points possess different upper bounds for Lagrange multipliers, it is usually required to locate the real outlier. The candidate

having the greatest frequency of suspected outliers having varying values of ε, is deemed as an outlier after numerous computations of SVR values *f(x)*. This technique is done repeatedly until no further outliers are identified or until the mean square error (training error) falls below the set threshold.

When this approach is applied to real-world occurrences, the following concerns arise. Firstly, in order to manage data with several outliers, the technique necessitates significant computation costs, as detecting an outlier requires repeated rounds of the optimization calculation. Secondly, non-expert users will find it difficult to use since it demands exact detection. Thirdly, SV method has a unique benefit based on SVM theory, in which it creates its formation (Chuang et al., 2002), signifying the chance of decreasing swamping and masking problems when utilizing various ε parameter values.

1.2.6 μ- ε-SVR Based Outlier Detection

Instead of using the parameter C, the μ- ε- SVR method for outlier detection uses parameter μ (new regularization parameter) introduced by Nakayama and Yun (2006) as a solution to solve the difficulties of the standard approach problems (Jordaan and Smits, 2004). Moreover, the μ- ε-SVR algorithm (Nishiguchi et al., 2010), was established in the following way:

$$\begin{aligned} &minimize && \frac{1}{2}\|w\|^2 + \mu\,(\xi_i + \xi_i^*) \\ &subject\ to && \begin{cases} y_i - w.\Phi(x_i) - b \le \varepsilon + \xi_i \\ w.\Phi(x_i) + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0, i = 1,2 \ldots., n. \end{cases} \end{aligned} \tag{1.13}$$

Only the highest training errors (ξ_i *or* ξ_i^*) are considered in the primary formation of the μ- ε-SVR method (1.13), not the average slack variables used in the standard SVR (Jordaan and Smits, 2004). From (1.13), the Lagrange function is constructed as follows:

$$\begin{aligned} L = \frac{1}{2}\|w\|^2 + \mu \sum_{i=1}^{n}(\xi_i + \xi_i^*) - \sum_{i=1}^{n}(\eta_i\xi_i + \eta_i^*\xi_i^*) \\ - \sum_{i=1}^{n}\alpha_i(\varepsilon + \xi_i - \quad y_i + w.\Phi(x_i) + b) \\ - \sum_{i=1}^{n}\alpha_i^*(\varepsilon + \xi_i^* + \quad y_i - w.\Phi(x_i) - b)\,. \end{aligned}$$

Based on the saddle point condition, we can see that the partial derivatives of function *L* with respect to the primary variables (w, b, ξ_i, ξ_i^*) provide the following dual optimization problem.

$$
\begin{aligned}
&maximize && \begin{cases} -\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)k(x_i,x_j) \\ -\varepsilon\sum_{i,j=1}^{n}(\alpha_i+\alpha_i^*)+\sum_{i=1}^{n}y_i(\alpha_i-\alpha_i^*) \end{cases} \\
&subject\ to && \sum_{i=1}^{n}(\alpha_i-\alpha_i^*)=0 \ \ and \ \ \alpha_i,\alpha_i^* \geq 0 \\
& && \sum_{i=1}^{n}\alpha_i \leq \mu, \sum_{i=1}^{n}\alpha_i^* \leq \mu
\end{aligned}
\tag{1.14}
$$

With regard to KKT conditions, the Lagrange multipliers sum is employed rather than their values by computing (1.14).

$$
\begin{aligned}
\xi_i\left(\mu-\sum_{i=1}^{n}\alpha_i\right)=0, \\
\xi_i^*\left(\mu-\sum_{i=1}^{n}\alpha_i^*\right)=0.
\end{aligned}
\tag{1.15}
$$

Each data points having non-zero (positive) Lagrange multipliers owns the same maximum error when the Lagrange multipliers sum (1.15) reaches the upper bound μ. Thus, these data points are most likely the actual outliers. Hence, provided that an outlier exists, the point having the greatest Lagrange multiplier is considered as the most probable outlier among points, according to the optimization concept. All Lagrange multipliers are not confined by the upper bound, as shown by the dual problem (1.14). The upper bounded Lagrange multiplier for several data points is always different. It is important to highlight that, any training errors less than the ε zone, or outliers, will not occur if the sum of the Lagrange multipliers is less than the upper bound μ. Hence, the outlier detection technique can employ the sum of the Lagrange multipliers. Outlier detection with the μ-ε-SVR algorithm is given below:

Step 1: Compute μ-ε –SVR.

Step 2: Determine the greatest α_i, α_i^*.

Step 3: From the data set, remove x_i and y_i with the highest α_i, α_i^*.

Step 4: Repeat the process until no more outliers are found.

When this approach is used in real-world applications, the following limitations occur. Firstly, because it can only discover and eliminate one outlier every iteration, this method is best for data with few outliers. When outliers are increased, computational costs approach those resulting from the standard SVR technique. Secondly, despite the fact

that it has a fixed tolerance ε, which tend to reduce swamping and masking issues, there is no clear criteria for determining the ε parameter value.

1.3 Basic Concepts

The robust regression approach was introduced with the goal of providing resistant estimates in the outliers' presence in the data set. Theoretically, the most basic qualities used to quantify the performance of resilient approaches are breakdown point and bounded influence. These robust estimator principles are briefly stated as the following.

1.3.1 Breakdown Point

A high breakdown point is another desired characteristic of a robust approach. The breakdown point (BP) is the minimum contamination percentage that can entirely ruin or implode an estimator or estimating process (Hampel, 1974, Coakley and Hettmansperger, 1993). Conversely, the smallest number of bad data (outliers) can drastically alter an estimator. In most cases, a high breakdown point indicates that the estimator can endure a significant number of outliers without the analysis imploding. Since the estimate remains bounded when fewer than 0.50 of the data are replaced with outlying observations, the largest attainable BP is 0.50 (Rousseeuw and Croux, 1993). Moreover, to propose a breakdown's formal finite sample definition, we may utilize a sample of n data point given as follows:

$$G = \{(x_{11}, \dots, x_{1p}, y_1), \dots, (x_{n1}, \dots, x_{np}, y_n)\}.$$

If we assume T to be a regression estimator, we get the following vector of regression coefficients when we apply T to such a sample G:

$$T(G) = \hat{\beta}.$$

To obtain all possible corrupted samples G^T, any m of the original data points is replaced with arbitrary values, or also known as outliers. Therefore, the estimator $T's$ breakdown point at sample G is defined as

$$BP(T, G) = \min\left\{\frac{m}{n}; \underset{G^T}{\text{SUP}} \|T(G^T) - T(G)\| \text{ is infinite}\right\}$$

in which the supremum is over all possible data matrix G, which contains $n - m$ observation and m contaminated points (Rousseeuw and Leroy, 1987; Maronna et al. 2006).

1.3.2 Bounded Influence Function

A robust for high leverage points or *X* space is shown by the bounded influence function (BIF) (Simpson, 1995). Specifically, BIF is an ability to protect the model estimators from the outlying points' effect in the *X* space. Moreover, the influence function (IF) assesses an estimator's robustness with regards to low contamination levels, and is frequently employed to determine either the estimator possesses BIF. The following is the IF of an estimator T at a distribution F in those points x_0 of the sample space when the limit exists:

$$\mathrm{IF}(x_0; T, F) = \lim_{\delta \to \infty} \frac{T((1-\delta)F + \delta\varphi_{x_0}) - T(F)}{\delta},$$

where φ_{x_0} denotes the probability distribution that puts all its mass in the point x_0 and δ represents the contamination amount. Moreover, it is vital to note that the influence function reflects the bias introduced by a few outliers at the point x_0 (Rousseeuw and Leroy, 1987; Simpson, 1995; Wilcox, 2005; Maronna, 2006).

1.4 Standardized

To limit the impacts of unit variation, the standardized form is often employed in numerous multiple regression applications. Therefore, the Standardized form is formulated this way for all $i = 1,2,\dots,n \; and \; j = 1,2,\dots,p$ (Kutner et al., 2005; Montgomery et al., 2015).

$$x_{ij}^{*} = \frac{x_{ij} - \bar{x}_j}{S_j},$$
$$y_i^{*} = \frac{y_i - \bar{y}}{S_y},$$

in which S_y and $\bar{y}$ represent the standard deviation and mean of the dependent variable y, respectively.

$$\bar{x}_j = \frac{\sum_{i=1}^{n} x_{ij}}{n}, \qquad S_j = \sqrt{\frac{\sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)^2}{n-1}}.$$
$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}, \qquad S_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}}.$$

1.5 Problem Statement

Several researchers have observed that true data sets include a high percentage of outliers spanning from 1% to 10% (Hampel et al., 1986; Wilcox, 2005). The existence of one or numerous outliers in a data set affects both parametric and nonparametric regression methods (the nonparametric methods are less swayed than the parametric methods). High leverage points (HLPs), as well as outliers, possess a significant impact on the various estimation values, resulting in inaccurate results and actions. Hence, it is very crucial to identify them before constructing a predictive model (Cook, 1977) or orienting resilient approaches (Huber, 1973). Moreover, these parametric approaches are unable to handle data that is less than complete. To overcome this issue, several academics turn to non-parametric methods to detect outliers in both full rank and less than full rank scenarios. For outlier detection in high dimensional data (HDD) that refer to a situation when $p >> n$, Jordaan and Smits (2004) proposed utilizing standard support vector regression (SSVR). This strategy works by repeatedly performing the SV regression model and detecting points that are thought to be outliers. When it comes to real-world applications, Nishiguchi et al. (2010) pointed out certain issues. Multiple outliers in the data demand significant computing costs because outlier detection necessitates a lot of rounds of the calculation; trial and error are employed for conclusive identifications because ways to find the outlier threshold value is unclear. To address the issue, Nishiguchi et al. (2010) established the modified support vector regression (MSVR) approach for outlier detection by adopting a new trade-off parameter (μ), where they are effective in detecting HLPs and outliers. However, because only one cycle is necessary to discover one outlier, the MSVR technique is acceptable for data with few outliers. As a result, in the presence of several outliers, computational costs are similar to those resulting from the traditional SVM regression approach. Furthermore, no strict precedent for determining the threshold parameter value, despite the fact that it has a fixed value.

To overcome those problems, Dhhan et al. (2015) introduced the fixed parameter support vector regression (FP-SVR). The FP-SVR approach shows good performance for detecting outliers from a single time iteration by using a fixed set of parameters. However, the FP-SVR performs well for $p < n$, but when applied to high dimensional data, it suffers from masking and swamping issues. As a results, the FP-SVR is not very successful in the detection of outliers under a variety of scenarios. Hence, their work has motivated us to develop a new method of identification of outliers in HDD which is expected to be very successful in the detection of outliers with the least percentage of masking and swamping effects. We call this method *nu*-support vector regression, denoted as *nu*-SVR.

Moreover, this thesis also focuses on robust approaches to heading the problem of the presence of HLPs in high dimensional data (HDD) in multiple linear regression models. As previously stated, the existence of outliers has a significant impact on the OLS estimator as well as it cannot be applied in HDD. There are several robust regression methods in the literature, that are alternatives to OLS, such as the least median of squares (LMS), least absolute values (LAV), the least trimmed squares (LTS), the maximal likelihood estimator (M-estimator), the method of moment estimator (MM estimator), the generalized M-estimator (GM1-estimator), and a new class of GM-estimator (GM6)

introduced by Coakley and Hettmansperger (1993). However, all these methods previously mentioned are not efficient enough in the presence of multiple outliers in the linear regression model, where these methods suffer from masking and swamping effects. To solve this problem, Dhhan et al. (2017) proposed a new version of the GM6 estimator based on SVR, which is called the GM-SVR. The GM-SVR technique performs well when predictors (p) are less than the sample size (n), but it cannot be applied when $p >> n$. To handle the dimensionality problems, Partial Least Squares Regression (PLSR) was developed to produce a regression model with multicollinear data in HDD. The aim is to use algorithms to extract uncorrelated latent variables (components) (Wold et al. (1983)). The statistically inspired modification of the PLS (SIMPLS) is one of the most popular algorithms for extracting such components iteratively (De Jong, 1993). Due to the use of the OLS estimation technique and a non-robust covariance matrix in collecting the components, it is now clear that the SIMPLS algorithm is very sensitive to outliers. Alin and Agostinelli (2017) introduced the Robust Iteratively Reweighted SIMPLS (RWSIMPLS) by using a weight function devised by Markatou et al. (1998), which is based on the response variable, the model distribution chosen, and the sample empirical distribution. However, using this weight function has shortcomings as there is no specific rule to indicate which observations are the outliers. This limitation has inspired us to develop another version of RWSIMPLS based on *nu*-SVR, denoted as SVR-RWSIMPLS.

This thesis is also concerned on constructing diagnostic plots to classify observations into four categories—regular observations, vertical outliers (outlying observations in Y-space), good leverage points (outlying observations in X space where they follow the pattern of the rest of the data) and bad leverage points (outlying observations in both X-space and Y-space) for HDD. Alin and Agostinelli (2017) and Alin et al. (2019) proposed a diagnostic plot for HDD by plotting the robust standardized residuals of the RWSIMPLS versus the leverage values. This turned out to be not a good approach since it is now evident that leverage values are not very successful in identifying HLPs (Habshah et al., 2009). This weaknesses has motivated us to develop a better approach of classifying observations into the four categories mentioned above.

This thesis also addresses the issue of variable selection in HDD. Some shrinkage procedures based on penalized function, such as the least absolute shrinkage and selection operator (LASSO) (Tibshirani, 1996), adaptive LASSO (Zou, 2006), and Elastic-Net (Zou and Hasti, 2005) were proposed to simultaneously perform coefficient estimation and variable selection for HDD. Unfortunately, all of these methods are not resistant to outliers and unable to select the importance variables. The LAD-LASSO (Xu, 2005; Wang and Leng, 2007) was put forward to address this issue. Nonetheless, the LAD-LASSO is still not very successful in selecting the important variables because it is only resistant to vertical outliers but not resistant to high leverage points. To remedy this problem, Arslan (2012) developed Weighted LAD-LASSO (WLAD-LASSO) by combining the Weighted LAD (WLAD) criterion with the L_1 penalty function. He utilized weight function which is obtained from robust mahalanobis distance (RMD) based on minimum volume ellipsoid (MVE). However, it is now evident that RMD-MVE suffers from swamping and masking effects (Dhhan et al. ,2015; Rashid et al. ,2021). Moreover, the WLAD-LASSO can only be applied to low dimensional data but

not for HDD. The shortcomings of WLAD-LASSO have inspired us to develop a new variable selection approach by incorporating a new weight function obtained from *nu*-SVR. We call this method the robust weighted LAD LASSO based on *nu*-SVR, denoted as WLL-SVR which is anticipated to be more accurate in selecting only the significant predictors in the model.

In this thesis, the issue of dimension reduction is also addressed. Principal component analysis (PCA) is a popular method for analyzing dimensional space in order to achieve dimension reduction, which can be accomplished by selecting the most components that can explain the variability in the data (Barnett and Lewis, 1994). Unfortunately, outliers have a bad effect on the PCA method, and this may lead to misleading conclusions. Hence, this drawback, has motivated us to develop a new approach that can achieve robustness and dimension reduction, known as the PCA-SVR method.

1.6 Research Objectives

This thesis's major purpose is to look at the challenges of high dimensionality in nonlinear and linear regression models when there are outliers (outlying in coordinates X and Y). The Ordinary least squares (OLS) estimates are used in most conventional diagnostic and estimation procedures for outliers' situations. Regrettably, the OLS estimate is not resistant to outliers. Furthermore, it is inappropriate for nonlinear regression models, and meeting all of its assumptions for high-dimensional models is challenging (full or less than full rank).

It is important to leave the classic models and the search for alternative that is flexible to be more resistant against outliers and appropriate for linear and nonlinear problems, whether it has a complete or incomplete rank. Our research's primary goals can be summarized in the following manner:

1. To develop a new detection method based on *nu*-SVR for the identification of high leverage points and vertical outliers in high dimensional data.
2. To establish a new robust partial least squares estimation method based on *nu*-SVR to remedy the presence of leverage points and outliers in high dimensional data.
3. To formulate a new classification scheme to classify observations into regular observations, vertical outliers, good and bad leverage points in high dimensional data.
4. To develop a robust WLAD-LASSO variable selection procedure in the presence of HLPs and correlated variables in high dimensional data.
5. To develop a new robust principal component analysis based on the *nu*-SVR model to overcome the curse of high dimensionality by achieving the dimensional reduction in the multiple linear regression model.

1.7 Study's Limitation and Scope

Many disciplines of study, e.g., bioinformatics, economics, financial forecasting, chemometrics, gene expressions and many more deal with high dimensional data. The analysis of high dimensional data become increasingly important in many fields and forms a major statistical challenge in terms of data classification and other statistical analysis. Therefore the scope of this thesis focusses on the development of several robust methods in high dimensional data.

Several problems arise when dealing with high dimensional data. The main problem is that a matrix related to some algorithms may become singular. Moreover, most of the classical procedures are easily affected by outliers and consequently the entire classical inferential procedures which rely on certain assumptions, especially the normality assumption, will give inaccurate predictions. Hence, the thesis also focusses on using robust method and nonparametric methods which do not depend on certain assumptions to handle those issues.

The nonparametric procedure refers to another statistical modelling technique that is used to examine high-dimensional and nonlinear relationships challenges. The SVM is among the most efficient algorithms in the nonparametric machine learning community (Frohlich and Zell, 2005). Nevertheless, when the threshold is low, the SVM model's capacity to evaluate high-dimensional issues is hindered since the resulting model is non-sparse.

The analysis of high dimensional data is computationally expensive and requires a large number of steps and long computer running times to complete. Since most of the analysis were done using lap-top computer (Intel(R) Core i5, 3ed generation, 2CUP**)** , it took several hours of computational running times to get certain results unless we can get access to working under high performance computer (HPC). Due to this problem, most work dealing with HDD, considered p up to 1,000 and percentage of contaminations only up to 20%. Hence, in certain chapters of this thesis, we are able to run the simulation study only up to p equals to 1,000. Moreover, the percentage of contamination is limited up to 20%.

1.8 Overview of the Thesis

This thesis's contents are organized into eight chapters in line with the study's goals and scope.

Chapter Two: The least squares estimation technique's literature review and violation of its basic assumptions (the outliers' existence and departure from normality) are covered in this section. The literature study of the support vector machine for regression is highlighted, as well as the core principle of using the kernel trick during the estimation

process. Leverage points and outliers are also explored, as well as their diagnostic approaches. Basic notions of robust linear regression are also discussed, as well as some major existing robust regression algorithms. The core principle of the partial least squares regression model, as well as its estimating methods, are explained in this section. The idea of variable selection is also discussed, as are certain penalization strategies. Lastly, the approach for principal component analysis and dimension reduction is also explored.

Chapter Three: This chapter is mainly devoted to developing a new method of identification of multiple HLPs in high dimensional data and named it *nu*-SVR. In this chapter, the performance of *nu*-SVR is evaluated using real data sets and simulation studies.

Chapter Four: The establishment of the RWSIMPLS estimation algorithm based on *nu*-SVR is presented in this chapter. Monte Carlo simulation studies and two real data sets are employed to assess the performance of the proposed method.

Chapter Five: In this chapter, a new variable selection method that we called WLL-SVR is discussed. The proposed method is evaluated through simulation studies and two real data sets.

Chapter Six: A hybrid principal component analysis and support vector regression (denoted by PCA-SVR) is developed in this chapter. A Monte Carlo simulation studies and real data sets are given to assess the performance of the proposed method.

Chapter Seven: The thesis conclusions are summarized and discussed in depth in this chapter, followed by suggestion for future study.

REFERENCES

Algamal, Z. Y., & Lee, M. H. (2015). Adjusted adaptive lasso in high-dimensional poisson regression model. *Modern Applied Science*, 9(4), pp170-177.

Alguraibawi, M., Midi, H., & Imon, A. H. M. (2015). A new robust diagnostic plot for classifying good and bad high leverage points in a multiple linear regression model. *Mathematical Problems in Engineering*, 2015.

Alin, A., & Agostinelli, C. (2017). Robust iteratively reweighted SIMPLS. *Journal of Chemometrics*, 31(3), e2881.

Alin, A., Agostinelli, C., Gergov, G., Katsarov, P., & Al-Degs, Y. (2019). Robust multivariate diagnostics for PLSR and application on high dimensional spectrally overlapped drug systems. *Journal of Statistical Computation and Simulation*, 89(6), 966-984.

Alter, O., Brown, P. O., & Botstein, D. (2000). Singular value decomposition for genome-wide expression data processing and modeling. *Proceedings of the National Academy of Sciences*, 97(18), 10101-10106.

Andersen, R. (2008). *Modern methods for robust regression* (No. 152). Sage.

Anderson, C., & Schumacker, R. E. (2003). A comparison of five robust regression methods with ordinary least squares regression: Relative efficiency, bias, and test of the null hypothesis. Understanding Statistics: Statistical Issues in Psychology, *Education, and the Social Sciences*, 2(2), 79-103.

Arslan, O. (2012). Weighted LAD-LASSO method for robust parameter estimation and variable selection in regression. *Computational Statistics & Data Analysis*, 56(6), 1952-1965.

Bagheri, A., & Midi, H. (2015). Diagnostic plot for the identification of high leverage collinearity-influential observations. *SORT-Statistics and Operations Research Transactions*, 51-70.

Balfer, J., & Bajorath, J. (2015). Systematic artifacts in support vector regression-based compound potency prediction revealed by statistical and activity landscape analysis. *PloS one*, 10(3), e0119301.

Barnett, V., & Lewis, T. (1994). *Outliers in statistical data*. Wiley. New York.

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*. John & Wiley, New York.

Ben-Hur, A., Ong, C. S., Sonnenburg, S., Schölkopf, B., & Rätsch, G. (2008). Support vector machines and kernels for computational biology. *PLoS Comput Biol*, 4(10), e1000173.

Bermingham, M., Pong-Wong, R., Spiliopoulou, A., Hayward, C., Rudan, I., Campbell, H., Navarro, P. (2015). Application of high-dimensional feature selection: Evaluation for genomic prediction in man. *Scientific Reports*, 5(10312), pp 1-12.

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, 144-152.

Boudt, K., Rousseeuw, P. J., Vanduffel, S., & Verdonck, T. (2020). The minimum regularized covariance determinant estimator. *Statistics and Computing*, 30(1), 113-128.

Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*, Cambridge University Press. Cambridge.

Branden, K. V., & Hubert, M. (2004). Robustness properties of a robust partial least squares regression method. *Analytica Chimica Acta*, 515(1), 229-241.

Breiman, L. (1996). Stacked regressions. *Machine learning*, 24(1), 49-64.

Bühlmann, P., & Van De Geer, S. (2011). *Statistics for high-dimensional data: methods, theory and applications.* Springer Science & Business Media.

Caner, M., & Fan, Q. (2010). The adaptive lasso method for instrumental variable selection. *Technical Report*.

Ceperic, V., Gielen, G., & Baric, A. (2014). Sparse ε-tube support vector regression by active learning. *Soft Computing*, 18(6), 1113-1126.

Chang, C. C., & Lin, C. J. (2002). Training v-support vector regression: theory and algorithms. *Neural computation*, 14(8), 1959-1977.

Chen, Y., Ma, J., Zhang, P., Liu, F., & Mei, S. (2015). Robust state estimator based on maximum exponential absolute value. *IEEE Transactions on Smart Grid*, 8(4), 1537-1544.

Chen, Y., Yao, Y., & Zhang, Y. (2020). A robust state estimation method based on SOCP for integrated electricity-heat system. *IEEE Transactions on Smart Grid*, 12(1), 810-820.

Cherkassky, V., & Mulier, F. M. (2007). *Learning from data: concepts, theory, and methods*. John Wiley & Sons.

Chuang, C., Su, S., Jeng, J., & Hsiao, C. (2002). Robust support vector regression networks for function approximation with outliers. *Neural Networks, IEEE Transactions on*, 13(6), pp 1322-1330.

Coakley, C. W., & Hettmansperger, T. P. (1993). A bounded influence, high breakdown, efficient regression estimator. *Journal of the American Statistical Association*, 88(423), 872-880.

Colliez, J., Dufrenois, F., & Hamad, D. (2006). Robust regression and outlier detection with SVR: Application to optic flow estimation. *The 17th British Machine Vision Association (BMVC)*, Edinburgh, UK, pp 1229-1238.

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, 19(1), pp 15-18.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273-297.

Cummins, D. J., & Andrews, C. W. (1995). Iteratively reweighted partial least squares: A performance analysis by Monte Carlo simulation. *Journal of Chemometrics*, 9(6), 489-507.

De Brabanter, K., De Brabanter, J., Suykens, J. A., & De Moor, B. (2010). Optimized fixed-size kernel models for large data sets. *Computational Statistics & Data Analysis*, 54(6), 1484-1504.

De Jong, S. (1993). SIMPLS: an alternative approach to partial least squares regression. *Chemometrics and intelligent laboratory systems*, 18(3), 251-263.

Dhhan, W., & Alshaybawee, T. (2017). Elastic net for single index support vector regression model. *Economic Computation & Economic Cybernetics Studies & Research*, 51(2).

Dhhan, W., Rana, S., & Midi, H. (2015). Non-sparse ϵ-insensitive support vector regression for outlier detection. *Journal of Applied Statistics*, 42(8), 1723-1739.

Dhhan, W., Rana, S., & Midi, H. (2017). A high breakdown, high efficiency and bounded influence modified GM estimator based on support vector regression. *Journal of Applied Statistics*, 44(4), 700-714.

Dhhan, W., Rana, S., Alshaybawee, T., & Midi, H. (2018). The single-index support vector regression model to address the problem of high dimensionality. *Communications in Statistics-Simulation and Computation*, 47(9), 2792-2799.

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, 32(2), 407-499.

Esbensen, K., Schoňkopf, S., & Midtgaard, T. (1995). Multivariate analysis in practice: Training package. *Computer-Aided Modelling*.

Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, 96(456), 1348-1360.

Fan, J., & Peng, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *The annals of statistics*, 32(3), 928-961.

Figueiredo, M. A. (2003). Adaptive sparseness for supervised learning. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(9), 1150-1159.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). Vol. 1 of *The elements of statistical learning.*

Frohlich, H., & Zell, A. (2005). Efficient parameter selection for support vector machines in classification and regression via model-based global optimization. Neural Networks, 2005. IJCNN'05. Proceedings. 2005 IEEE International Joint Conference on, 3 1431-1436.

Furno, M. (1996). Small sample behavior of a robust heteroskedasticity consistent covariance matrix estimator. *Journal of Statistical Computation and Simulation*, 54(1-3), 115-128.

Gan, L., Lv, W., Zhang, X., & Meng, X. (2012). Improved PCA+ LDA applies to gastric cancer image classification process. *Physics Procedia*, 24, 1689-1695.

Groß, J. (2003). *Linear regression*. Springer, Heidelberg, Germany.

Guo, B., Gunn, S. R., Damper, R. I., & Nelson, J. D. (2008). Customizing kernel functions for SVM-based hyperspectral image classification. *Image Processing, IEEE Transactions on*, 17(4), 622-629.

Guo, G., Zhang, J., & Zhang, G. (2010). A method to sparsify the solution of support vector regression. *Neural Computing and Applications*, 19(1), 115-122.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *The Journal of Machine Learning Research*, 3, 1157-1182.

Habshah, M., Norazan, M. R., & Rahmatullah Imon, A. H. M. (2009). The performance of diagnostic-robust generalized potentials for the identification of multiple high leverage points in linear regression. *Journal of Applied Statistics*, 36(5), 507-520.

Hadi, A. S. (1992). A new measure of overall potential influence in linear regression. *Computational Statistics & Data Analysis*, 14(1), 1-27.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69(346), 383-393.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (2011). *Robust statistics: the approach based on influence functions* (Vol. 196). John Wiley & Sons.

Hampel, F., Ronchetti, E., Rousseeuw, P., & Stahel, W. (1986). *Robust statistics*, J. Wiley& Sons, New York.

Härdle, W. K., Hoffmann, L., & Moro, R. (2011). Learning machines supporting bankruptcy prediction. *Statistical tools for finance and insurance*. Springer, Heidelberg, pp. 225-250.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). Unsupervised learning. *The elements of statistical learning* (pp. 485-585).

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2, pp. 1-758). New York: springer.

Hawkins, D. M. (1980). *Identification of outliers*. Chapman and Hall, London.

Hoerl, A. E., & Kennard, R. W. (1970a). Ridge regression: Applications to nonorthogonal problems. *Technometrics*, 12(1), 69-82.

Hoerl, A. E., & Kennard, R. W. (1970b). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55-67.

Horowitz, J. L. (2009). *Semiparametric and nonparametric methods in econometrics*. Springer, New York.

Huang, X., Wu, L., & Ye, Y. (2019). A review on dimensionality reduction techniques. *International Journal of Pattern Recognition and Artificial Intelligence*, 33(10), 1950017.

Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1), pp 73-101.

Huber, P. J. (1973). Robust regression: Asymptotics, conjectures and monte carlo. *The Annals of Statistics*,1(5), pp 799-821.

Hubert, M., & Branden, K. V. (2003). Robust methods for partial least squares regression. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 17(10), 537-549.

Hubert, M., Rousseeuw, P. J., & Vanden Branden, K. (2005). ROBPCA: a new approach to robust principal component analysis. *Technometrics*, 47(1), 64-79.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. Journal of educational psychology, 24(6), 417.

Imon, A. H. M. R. (2002). Identifying multiple high leverage points in linear regression. *Journal of Statistical Studies*, 3, 207-218.

Imon, A.H.M.R. (2005). Identifying multiple influential observations in linear regression. Journal of Applied Statistics. 32: 929-946.

Ismaeel, S. S., Midi, H., & Sani, M. (2021). Robust Multicollinearity Diagnostic Measure For Fixed Effect Panel Data Model. Ma*laysian Journal of Fundamental and Applied Sciences*, 17(5), 636-646.

Jordaan, E. M., & Smits, G. F. (2004). Robust outlier detection using SVM regression. Neural Networks, 2004. Proceedings. *2004 IEEE International Joint Conference on*, 3 2017-2022.

Kalivas, J. H. (1997). Two data sets of near infrared spectra. *Chemometrics and Intelligent Laboratory Systems*, 37(2), 255-259.

Kamruzzaman, M., & Imon, A. (2002). High leverage point: Another source of multicollinearity. *Pakistan Journal of Statistics-All Series-*, *18*(3), pp 435-448.

Karatzoglou, A., Meyer, D., & Hornik, K. (2006). Support vector machines in R. *Journal of statistical software*, 15(1), 1-28.

Karatzoglou, A., Smola, A., Hornik, K., & Zeileis, A. (2004). kernlab-an S4 package for kernel methods in R. *Journal of statistical software*, 11(9), 1-20.

Keerthi, S. S., Shevade, S. K., Bhattacharyya, C., & Murthy, K. R. K. (2001). Improvements to platt's SMO algorithm for SVM classifier design. *Neural Computation*, 13(3), 637-649.

Kuo, T., & Yajima, Y. (2010). Ranking and selecting terms for text categorization via SVM discriminate boundary. *International Journal of Intelligent Systems*, 25(2), 137-154.

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). Applied linear statistical models. McGraw-Hill Irwin, New York.

Lahiri, S. K., & Ghanta, K. C. (2009). Support vector regression with parameter tuning assisted by differential evolution technique: Study on pressure drop of slurry flow in pipeline. *Korean journal of chemical engineering*, 26(5), 1175-1185.

Lee, Y., & Mangasarian, O. L. (2001). SSVM: A smooth support vector machine for classification. *Computational Optimization and Applications*, 20(1), 5-22.

Lim, H. A., & Midi, H. (2016). Diagnostic Robust Generalized Potential Based on Index Set Equality (DRGP (ISE)) for the identification of high leverage points in linear model. *Computational Statistics*, 31(3), 859-877.

Liu, J. N., & Hu, Y. (2013). Support vector regression with kernel mahalanobis measure for financial forecast: In Time series analysis, *modeling and applications.* Springer, Heidelberg, pp. 215-227.

Markatou, M. (1996). Robust statistical inference: Weighted likelihoods or usual M-estimation. *Communications in Statistics-Theory and Methods*, 25(11), 2597-2613.

Markatou, M., Basu, A., & Lindsay, B. G. (1998). Weighted likelihood equations with bootstrap root search. *Journal of the American Statistical Association*, 93(442), 740-750.

Maronna, R. A., Martin, R. D., & Yohai, V. J. (2006). *Robust statistics*. John Wiley Chichester.

Maronna, R. A., Martin, R. D., Yohai, V. J., & Salibián-Barrera, M. (2019). *Robust statistics: theory and methods (with R)*. John Wiley & Sons.

Midi, H., Ismaeel, S. S., Arasan, J., & AMohammed, M. O. H. A. M. M. E. D. (2021). Simple and Fast Generalized-M (GM) Estimator and Its Application to Real Data Set. *Sains Malaysiana*, 50(3), 859-867.

Midi, H., Sani, M., Ismaeel, S. S., & Arasan, J. (2021). Fast Improvised Influential Distance for the Identification of Influential Observations in Multiple Linear Regression. *Sains Malaysiana*, 50(7), 2085-2094.

Midi, H., Sani, M., Ismaeel, S. S., & Arasan, J. (2021). Fast Improvised Influential Distance for the Identification of Influential Observations in Multiple Linear Regression. Sains Malaysiana, 50(7), 2085-2094.

Midi, H., & Mohammed, M. A. (2014). The Performance of Robust Latent Root Regression Based on MM and modified GM estimators. *WSEAS Transactions on Mathematics*, 13.

Mohammed Rashid, A., Midi, H., Dhhan, W., & Arasan, J. (2021). Detection of outliers in high-dimensional data using nu-support vector regression. *Journal of Applied Statistics*, 1-20.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2015). *Introduction to linear regression analysis*, John Wiley & Sons.

Muñoz-Garcia, J., Moreno-Rebollo, J., & Pascual-Acosta, A. (1990). Outliers: A formal approach. *International Statistical Review/Revue Internationale De Statistique*, *58*(3), pp 215-226.

Nakayama, H., & Yun, Y. (2006). Support vector regression based on goal programming and multi-objective programming. *In the 2006 IEEE International Joint Conference on Neural Network Proceedings*, Vancouver, BC, Canada.

Narasimhan, S., & Shah, S. L. (2008). Model identification and error covariance matrix estimation from noisy data using PCA. *Control Engineering Practice*, 16(1), 146-155.

Nishiguchi, J., Kaseda, C., Nakayama, H., Arakawa, M., & Yun, Y. (2010). Modified support vector regression in outlier detection. Neural Networks (IJCNN), the 2010 International Joint Conference on, 1-5.

Pell, R. J. (2000). Multiple outlier detection for multivariate calibration using robust statistical techniques. *Chemometrics and Intelligent Laboratory Systems*, 52(1), 87-104.

Rahmatullah Imon, A. H. M. (2005). Identifying multiple influential observations in linear regression. *Journal of Applied statistics*, 32(9), 929-946.

Rana, S., Dhhan, W., & Midi, H. (2018). Fixed parameters support vector regression for outlier detection. *Economic Computation & Economic Cybernetics Studies & Research*, 52(2).

Rana, S., Siraj-Ud-Doulah, M., Midi, H., & Imon, A. H. M. R. (2012). Decile mean: A new robust measure of central tendency. *Chiang Mai journal of science*, 39(3), 478-485.

Rashid, A. M., Midi, H., Slwabi, W. D., & Arasan, J. (2021). An Efficient Estimation and Classification Methods for High Dimensional Data Using Robust Iteratively Reweighted SIMPLS Algorithm Based on Nu-Support Vector Regression. *IEEE Access*, 9, 45955-45967.

RCore, T. E. A. M. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rojo-Álvarez, J. L., Martínez-Ramón, M., Figueiras-Vidal, A. R., García-Armada, A., & Artés-Rodríguez, A. (2003). A robust support vector algorithm for nonparametric spectral analysis. *IEEE Signal Processing Letters* 10(11), pp 320-323.

Roth, V. (2004). The generalized LASSO. *Neural Networks, IEEE Transactions on*, 15(1), 16-28.

Rousseeuw, P. J. (1985). Multivariate estimation with high breakdown point. *Mathematical Statistics and Applications*, 8, 283-297.

Rousseeuw, P. J., & Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, 88(424), 1273-1283.

Rousseeuw, P. J., & Leroy, A. M. (1987). *Robust regression and outlier detection*. New York: John Wiley & Sons.

Rousseeuw, P., & Van Zomeren, B. (1990). Unmasking multivariate outliers and leverage points (with discussion). *J.Amer.Statist.Assoc*, 85, 633-651.

Schölkopf, B., & Smola, A. (2002). *Learning with kernels*, MIT Press, Boston.

Schölkopf, B., Bartlett, P. L., Smola, A. J., & Williamson, R. (1999). Shrinking the tube: a new support vector regression algorithm. *Advances in neural information processing systems*, 330-336.

Schölkopf, B., Smola, A. J., Williamson, R. C., & Bartlett, P. L. (2000). New support vector algorithms. *Neural computation*, 12(5), 1207-1245.

Segal, M. R., Dahlquist, K. D., & Conklin, B. R. (2003). Regression approaches for microarray data analysis. *Journal of Computational Biology*, 10(6), 961-980.

Serneels, S., Croux, C., Filzmoser, P., & Van Espen, P. J. (2005). Partial robust M-regression. *Chemometrics and Intelligent Laboratory Systems*, 79(1-2), 55-64.

Simpson, J. R. (1995). New methods and comparative evaluations for robust and biased-robust regression estimation, unpublished PhD thesis, Arizona State University.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3), 199-222.

Smola, A., & Vapnik, V. (1997). Support vector regression machines. *Advances in Neural Information Processing Systems*, 9, 155-161.

Stamey, T. A., Kabalin, J. N., McNeal, J. E., Johnstone, I. M., Freiha, F., Redwine, E. A., & Yang, N. (1989). Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate. II. Radical prostatectomy treated patients. *The Journal of urology*, 141(5), 1076-1083.

Suykens, J. A., De Brabanter, J., Lukas, L., & Vandewalle, J. (2002). Weighted least squares support vector machines: robustness and sparse approximation. *Neurocomputing*, 48(1-4), 85-105.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288.

Tipping, M. E. (2001). Sparse bayesian learning and the relevance vector machine. *The Journal of Machine Learning Research*, 1, 211-244.

Ukil, A. (2007). *Intelligent systems and signal processing in power engineering*, Springer, Heidelberg.

Uraibi, H., & Midi, H. (2020). Robust Variable Selection Method Based on Huberized Lars-Lasso Regression. *Economic Computation & Economic Cybernetics Studies & Research*, 54(3).

Üstün, B., Melssen, W. J., & Buydens, L. M. (2006). Facilitating the application of support vector regression by using a universal Pearson VII function based kernel. *Chemometrics and Intelligent Laboratory Systems*, 81(1), 29-40.

Üstün, B., Melssen, W., Oudenhuijzen, M., & Buydens, L. (2005). Determination of optimal support vector regression parameters by genetic algorithms and simplex optimization. *Analytica Chimica Acta*, 544(1), 292-305.

Van Der Maaten, L., Postma, E., & Van den Herik, J. (2009). Dimensionality reduction: a comparative. *J Mach Learn Res*, 10(66-71), 13.

Vanderbei, R. J. (1999). LOQO user's manual—version 3.10. *Optimization Methods and Software*, 11(1-4), 485-514.

Vapnik, V. (1995). *The nature of statistical learning theory*, 1st ed. Springer, New York.

Vapnik, V. (1999). *The nature of statistical learning theory*. Springer science & business media.

Vapnik, V., Golowich, S. E., & Smola, A. (1996). Support vector method for function approximation, regression estimation, and signal processing. *Advances in Neural Information Processing Systems*, vol 9. MIT Press, p 281-287.

Varmuza, K., & Filzmoser, P. (2016). Introduction to multivariate statistical analysis in chemometrics. CRC press.

Velleman, P. F., & Welsch, R. E. (1981). Efficient computing of regression diagnostics. *The American Statistician*, 35(4), 234-242.

Virmani, J., Dey, N., & Kumar, V. (2016). PCA-PNN and PCA-SVM based CAD systems for breast density classification. In Applications of intelligent optimization in biology and medicine (pp. 159-180). Springer, Cham.

Wahid, A., Khan, D. M., & Hussain, I. (2017). Robust Adaptive Lasso method for parameter's estimation and variable selection in high-dimensional sparse models. *PLoS one*, 12(8), e0183518.

Wakelinc, I. N., & Macfie, H. J. H. (1992). A robust PLS procedure. *Journal of Chemometrics*, 6(4), 189-198.

Wang, H., & Leng, C. (2007). Unified LASSO estimation by least squares approximation. Journal of the American Statistical Association, 102(479), 1039-1048.

Wang, L., Cheng, H., Liu, Z., & Zhu, C. (2014). A robust elastic net approach for feature learning. *Journal of Visual Communication and Image Representation*, 25(2), 313-321.

Wang, T., & Li, Z. (2017). Outlier detection in high-dimensional regression model. *Communications in Statistics-Theory and Methods*, 46(14), 6947-6958.

Weisberg, S. (2005). *Applied linear regression*, John Wiley & Sons, Hoboken New Jersey.

Wilcox Rand, R. (2005). Introduction to robust estimation and hypothesis testing, *Elsevier academic Press*, New York.

Williams, G. (2011). *Data mining with rattle and R: The art of excavating data for knowledge discovery*, Springer, New York.

Wilson, H. G. (1978). Least squares versus minimum absolute deviations estimation in linear models. *Decision Sciences*, 9(2), 322-335.

Wold, S., Martens, H., & Wold, H. (1983). The multivariate calibration problem in chemistry solved by the PLS method. *In Matrix pencils* (pp. 286-293). Springer, Berlin, Heidelberg.

Xiang, L., Quanyin, Z., & Liuyang, W. (2013). Research of bessel kernel function of the first kind for support vector regression. *Information Technology Journal*, 12(14), 2673.

Xu, J. (2005). Parameter estimation, model selection and inferences in L (1)-based linear regression. Columbia University.

Yatchew, A. (2003). *Semiparametric regression for the applied econometrician Cambridge University Press*, Cambridge.

Ye, W., & Peng, C. (2018, October). Recognition algorithm of emitter signals based on PCA+ CNN. In 2018 IEEE 3rd Advanced Information Technology, *Electronic and Automation Control Conference* (IAEAC) (pp. 2410-2414). IEEE.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476), 1418-1429.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, 67(2), 301-320.

Zou, H., & Xue, L. (2018). A selective overview of sparse principal component analysis. *Proceedings of the IEEE*, 106(8), 1311-1320.

Zou, H., & Zhang, H. H. (2009). On the adaptive elastic-net with a diverging number of parameters. *Annals of statistics*, 37(4), 1733.