

**A NEW APPROACH TO
SUPPORT VECTOR REGRESSION (SVR) WITH INTERVAL
DATA**

By
SHIBSHANKAR DEY
Student ID: 0417082009 P

A thesis
Submitted to the
Department of Industrial and Production Engineering,
Bangladesh University of Engineering and Technology,
In partial fulfillment of the requirements
For the degree of
MASTER OF SCIENCE
In
Industrial and Production Engineering








November 2019

**DEPARTMENT OF INDUSTRIAL AND PRODUCTION ENGINEERING
BANGLADESH UNIVERSITY OF ENGINEERING AND TECHNOLOGY
DHAKA – 1000, BANGLADESH**

CERTIFICATE OF APPROVAL

The thesis titled “**A NEW APPROACH TO SUPPORT VECTOR REGRESSION (SVR) WITH INTRVAL DATA**” submitted by Shibshankar Dey, Student No. 0417082009P, Session: April – 2017, has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Master of Science in Industrial and Production Engineering on November 24, 2019.

BOARD OF EXAMINERS

 _____	
Dr. AKM Kais Bin Zaman Professor Department of IPE, BUET, Dhaka	Chairman (Supervisor)
 _____	
Dr. AKM Kais Bin Zaman Head Department of IPE, BUET, Dhaka	Member (Ex-Officio)
 _____	
Dr. Abdullahil Azeem Professor Department of IPE, BUET, Dhaka	Member
 _____	
Dr. Shuva Ghosh Assistant Professor Department of IPE, BUET, Dhaka	Member
 _____	
Dr. Md. Farhad Hossain Professor Department of EEE, BUET, Dhaka	Member (External)

CANDIDATE'S DECLARATION

It is hereby declared that this thesis or any part of it has not been submitted elsewhere for the award of any degree or diploma.



Shibshankar Dey

Student ID: 0417082009

ABSTRACT

This thesis develops several novel approaches to pursue the support vector regression (SVR) with interval data. Four approaches are proposed: (i) moment-based approach, (ii) equiprobability-based approach, (iii) boundary-point-based approach, and (iv) extended generalized-SVR. Applicability of each of the four proposed approaches varies depending upon the presence of the interval data in input and output observations. Interval data may be present only in input observations, or output observations, or in both. Thus, three cases may arise regarding the presence of interval data in input and output observations. All these three cases are considered in this thesis. The first two proposed approaches are applicable to all three cases. The boundary-point-based approach is applicable for the presence of interval data in output observations. The extended generalized-SVR approach is developed discerning the limitations of the existing generalized-SVR for the presence of interval data in both input and output observations. Therefore, extended generalized-SVR approach is applicable only when the interval data are present in both input and output observations. The separation strategy – where interval-valued inputs and outputs are dealt separately – is introduced to make the most time-consuming moment-based approach computationally tractable for the third case. This strategy is also utilized in proposing the extended generalized-SVR approach. The prediction accuracies and computational time of all the proposed approaches are compared within themselves as well as with the available existing method. Three real datasets and one synthetic dataset are used to experiment with the proposed approaches for their prediction accuracies and computational efficiency. Boundary-point-based approach and extended generalized-SVR approach are discerned as more efficient compared to the moment-based approach and equiprobability-based approach. It is shown that the moment-based approach always outperforms the equiprobability-based approach in terms of prediction accuracy for all three cases. However, prediction accuracy of the boundary-point based approach may be greater or less than that of the moment-based and equiprobability-based approach based on different cases. Prediction accuracies of existing generalized-SVR approach and extended generalized-SVR approach is always observed to be less than the other three approaches.

ACKNOWLEDGEMENT

By the grace of the most benevolent and Almighty God, the thesis titled “A New Approach to Support Vector Regression (SVR) with Interval Data” has been done and the report has been completed. The author of this report thinks that it is his moral duty to convey his earnest gratitude to a number of extra-ordinary people without whom this study would have not been possible.

Firstly, the author of this thesis report would like to express his sincere respect and heart-felt gratitude to his thesis supervisor Dr. AKM Kais Bin Zaman, Professor and Head, Department of Industrial and Production Engineering (IPE), Bangladesh University of Engineering and Technology (BUET), for his whole-hearted supervision. His clear guidance, timely instructions, invaluable advice, prudent suggestions, and sheer encouragements throughout the progress of the thesis and report writing have made this thesis possible.

Next, the author would like to convey his sincere gratitude to Dr. Abdullahil Azeem, Professor, Department of IPE, BUET, Dr. Shuva Ghosh, Assistant Professor, Department of IPE, BUET, and Dr. Md. Farhad Hossain, Professor, Department of Electrical and Electronic Engineering, BUET, for their constructive remarks and kind evaluations of this study.

The author would also like to acknowledge the kind assistance and whole-hearted support of Dr. AKM Kais Bin Zaman, Professor, Department of Industrial and Production Engineering (IPE), BUET for giving permission to use the resources of CAD Lab of IPE Department without which most of the computationally expensive simulation runs required during this study would be quite difficult and immensely time-consuming for the author.

The authors would like to express thanks and appreciations to all of his well-wishers who inspired him to continue this work. He is grateful to his family members for their support and encouragement. Last but not least, the author wants to convey his appreciations to all of those who have supported him in any respect during the study.

TABLE OF CONTENTS

ABSTRACT.....	4
ACKNOWLEDGEMENT	5
LIST OF TABLES.....	8
LIST OF FIGURES	9
CHAPTER 1	1
INTRODUCTION	1
1.1 Background of the Study	1
1.2 Contributions of the Present Study	3
1.3 Organization of the Thesis Report	5
CHAPTER 2	6
LITERATURE REVIEW	6
2.1 Classification Analysis Considering Uncertainty	6
2.2 Regression Analysis with Noisy Data.....	8
2.3 Regression Analysis with Interval Data.....	9
CHAPTER 3	1
BACKGROUND CONCEPTS REVIEW	1
3.1 Support Vector Regression	1
3.2 Kernels	9
3.3 Nonlinear SVR in Primal Form	11

3.4 Probabilistic approach for dealing with interval data	13
CHAPTER 4	17
PROPOSED METHODOLOGIES	17
4.1 Moment-based Approach.....	17
4.2 Equiprobability-based Approach	25
4.3 Boundary-point-based Approach	27
4.4 Extended generalized-SVR approach	31
CHAPTER 5	35
NUMERICAL EXPERIMENTATION	35
5.1 Case 1: Interval Data Present in Input Observations Only	35
5.2 Case 2: Interval Data Present in Output Observations only	39
5.3 Case 3: Interval Data Present in Both Input and Output Observations.....	45
5.4 Discussion of Findings.....	49
CHAPTER 6	51
CONCLUSIONS AND FUTURE SCOPES.....	51
6.1 Conclusions.....	51
6.2 Future Scopes.....	52
REFERENCES	54

LIST OF TABLES

Table 3.1 Common loss functions and corresponding density models.....	3
Table 3.2 Methods for calculating moment bounds for single interval data	15
Table 5.1 Prediction errors of moment-based approach and equiprobability-based approach during minimization and maximization in the outer loop	37
Table 5.2 Prediction errors for the concrete slump dataset using different approaches	37
Table 5.3 Computational time of different approaches for the concrete slump dataset	38
Table 5.4 Prediction errors for the unreliable sensor problem using different approaches	42
Table 5.5 Computational time of different approaches for the unreliable sensor problem	42
Table 5.6 Prediction errors for the wine quality dataset using different approaches.....	43
Table 5.7 Computational time of different approaches for the wine quality dataset.....	44
Table 5.8 Prediction errors from all the proposed approaches for the social survey dataset.....	46
Table 5.9 Computational time of different approaches for the wine quality dataset.....	49

LIST OF FIGURES

Figure 3.1 Soft margin loss setting for linear SVM in regression	5
Figure 3.2 Identification of Johnson distribution family	16
Figure 4.1 Pseudocode for the moment-based approach	20
Figure 4.2 Pseudocode for the moment-based approach with separation strategy	24
Figure 4.3 Pseudocode for the equiprobability-based approach	26
Figure 4.4 Pseudocode of Boundary-point-based approach	30
Figure 4.5 Pseudocode of extended generalized-SVR approach	34
Figure 5.1 Prediction bounds from the boundary-point-based approach for the wine quality data set	45

CHAPTER 1

INTRODUCTION

1.1 Background of the Study

Classification and Regression analysis are two main fields of machine learning (ML). Classification represents differentiating two or multiple classes based on the characteristic nature of the corresponding classes. If the number of classes to be differentiated is two, the classification problem is called binary classification, and while it is more than two, it is called multiclass classification. In binary classification, classes are represented generally by +1 and -1. In multiclass classification, various favorable categories are assigned to the classes for the convenience of identification. In short, the binary or multiclass classification problem of prediction stands for distinguishing the competitive categorical classes of response variables. However, when prediction problem contains non-categorical variables as response, regression analysis comes into action instead of classification. In other words, regression analysis stands for establishing a functional relationship between output and input variables where the output is generally a non-categorical variable while the inputs may be categorical or numerical. In classification and regression problems, outputs or responses are called dependent variables as outputs or responses are dependent on the inputs for being predicted through the learned relationship. On the other hand, since, in the prediction problem, input variables help to predict or explain the functional relationship between input and output, input variables are also defined as predictors or explanatory variables. A broad literature is present regarding regression or classification; however, when it appears uncertainty and impreciseness in data during prediction, the corresponding number of literature reduces to a great extent.

Impreciseness in data may result in interval data which arise due to other different reasons. Data rounding, data heaping, measurement instruments uncertain readings, data censoring are some of the common sources of interval data. Data censoring arises when failure does not happen at the time of observation but any time between the point of two intermittent observations while digit preference phenomenon best describes data heaping (Heitjan and Rubin, 1991). The modern era has a pronounced concern for security, privacy, competition etc. that results in restriction on

data collection, information gap, and intentional data obscuring which in turn evoke data collector to group the possible values of data. Such grouping yield in data binning - another form of interval data - which is also common in survey questionnaire, census information collection, professionals' opinion collection etc. (Ferson et al., 2007). When a subject gives multiple measures of the same quantity of interest, summarizing the measures yields interval data as common for measurement instruments uncertain readings. In the same manner, interval data arises while large unmanageable datasets are summarized through retaining enough knowledge from the original data (Carrizosa et al., 2007). Use of interval data is also a common phenomenon in material purity assessment and chemical quantification, where detection point may happen anywhere within the detection margin. In no way that detection point should be said to hold a certain likelihood value within that detection range as it is for the other mentioned sources of interval data. Thus, the main challenge with the interval data is to find the competitive point data from the respective intervals considering different types of uncertainty inherent in those interval data as dealt in this thesis.

There exist various methods for classification and regression analysis in ML field; however, uses of support vector machine (SVM) for classification and regression are getting the most floors in the recent time where such uses are known as support vector classification (SVC) and support vector regression (SVR), respectively (Vapnik, 2013; Vapnik et al., 1997; Drucker et al., 1997; Vapnik and Vapnik, 1998; Gunn, S.R., 1998; Schölkopf et al., 2002; Hsu et al., 2003; Smola and Schölkopf, 2004; Basak et al., 2007). However, in SVM, since only a part of its training data known as support vectors (SVs) is used for faster output prediction of a new test point, decision hyperplane easily becomes affected for contamination by aleatory uncertainty (i.e., feature noise) in SVs, which ultimately leads to poor accuracy in prediction. Moreover, the assumption of known probability distribution of the random noise, working with the midpoints of the interval data, etc. in the presence of epistemic uncertainty (i.e., imprecise probabilistic information due to sparse and interval data) further deteriorate SVM prediction accuracy.

To deal with the noisy data in SVC problem, besides the general statistical framework proposed by Bi and Zhang. (2005), use of pinball loss function (Huang et al., 2013; Xu et al., 2016) is also common. To keep the prediction model immune to the potential noises in input data, the robust approach proposed by Trafalis and Gilbert (2006) is applicable for both classification and regression. Generalizing the formulations proposed by the Trafalis and Gilbert (2006), a robust

classification model to deal with both noisy and interval-valued data is proposed by Carrizosa et al. (2007). In classification problem with the presence of noisy and interval data, alongside of SVM, uses of neural network (Rossi and Conan-Guez, 2002) and Kernel-based approaches (Do and Poulet, 2005) are also mention-worthy. However, in SVR problems, literature considering the noisy and imprecise data becomes more confined, especially, if we search for robust approach with interval data. Interval regression analysis is performed by Tanaka and Lee (1998), Hwang et al. (2006), Hao (2009) considering noisy data. Non-parametric regression analysis considering noisy data (Petit-Renaud and Dencœux, 2004), robust regression approach considering imprecise data (Cattaneo and Wiencierz, 2012; Cattaneo and Wiencierz, 2014) are some of the few available literature that pursue regression analysis with noisy and imprecise data, where SVM is not used for regression. To mention the very few literature using SVM for regression in presence of interval data, Utkin and Coolen (2011) can be cited. However, the generalized-SVR formulations given by Utkin and Coolen (2011) for the interval data do not consider the various types of uncertainties inherent in interval data, and also are not applicable for interval data present in both input and output observations. Thus, the presence of very few literature of SVR with interval-valued observations and the limitations of the generalized-SVR method give us the motivations to pursue this research. Therefore, in this thesis, we propose several approaches that perform the support vector regression (SVR) considering the uncertainty present in the interval data.

1.2 Contributions of the Present Study

The main contributions of this research are to develop different approaches in pursuing SVR with interval data. In proposing the approaches, all three different cases that arise from the presence of interval data are considered. These three cases are (i) interval data present only in input observations, (ii) interval data present only in output observations, and (iii) interval data present in both input and output observations. Note that the presence of interval data in input, or in output, or in both does not necessarily mean that all the corresponding observations are interval-valued. In other words, point data may also be present with the interval-valued observations. Such issue of coexistence of point data with the interval data is also considered in the proposed approaches. Four approaches are proposed - three of them are novel and the rest one is an extended version of the existing generalized-SVR approach. Four proposed approaches are equiprobability-based approach, moment-based approach, and boundary-points-based approach and extended

generalized-SVR approach. Although the first two approaches can be applied for all the three mentioned cases, the third approach is applicable only when interval-valued observations are present in output. The fourth approach actually extends the generalized-SVR framework to make it applicable for the presence of interval data in both input and output observations. When there are interval data in observed outputs (i.e., the second and the third case), all the proposed approaches are designed to predict an interval where the unobserved output can lie. For the first case, point prediction is performed with the first two proposed approaches. Prediction accuracies and computational time of all the proposed approaches are intercompared as well as compared with the existing generalized SVR method for the first two cases. For the third case, prediction accuracy and computational time are compared among all the proposed approaches.

Prediction is required in any engineering discipline. Prediction facilitates decision-making, which is one of the major areas of study in Industrial and Production Engineering (IPE). Accuracy and efficiency are the main concerns in any prediction problem as this thesis highlights. Thus, this research undoubtedly serves an important part in IPE discipline. Besides, if we look for the specific sources of interval data in IPE discipline, we would easily get hundreds of sources. As mentioned earlier, data heaping, data binning, data rounding, data censoring, and measurement instrument uncertainty are some of the many sources of interval data. Like any other engineering discipline, these sources are also common in IPE discipline. For example, a division of IPE is production or manufacturing engineering. To bring a new product in market, it needs to undergo many stages. Ensuring the physical features and performance of the manufactured products according to their design is one of the most challenging stages. To tackle such challenge, one of the practices is to perform accelerated life testing (ALT) for monitoring the health of the components or products at their different life stages. Some monitoring plans call for periodic or intermittent measurements. A common example of this is the regular inspection of the components. If a component is observed to be in good working condition at one inspection, but not at the next inspection, it cannot be precisely said when the component has failed. It seems entirely reasonable to conclude that there is a window of time between the last two inspections during which the component failed and the natural mathematical representation of that failure time is an interval. Moreover, uncertainty in the measurement devices used for inspection itself can yield interval data. The situation just described from the perspective of the life stages of manufactured products is just one of the many examples regarding the emergence of the interval data in IPE discipline. Hence, the contribution of this thesis

is to develop several prediction approaches in the presence of interval data, which facilitates solving real world engineering problems including problems in Industrial and Production Engineering.

1.3 Organization of the Thesis Report

The rest of the thesis paper is organized as follows. Chapter 2 represents a detail literature review in the field of prediction with noisy and interval data. Chapter 3 reviews the fundamental background concepts used in the proposed methods. Chapter 4 presents the detailed framework of all the proposed approaches. In Chapter 5, we examine the accuracy and computational efficiency of the proposed methods using four well-known datasets (concrete slump test dataset, unreliable sensor problem dataset, wine quality dataset, and German general social survey dataset) for all three different cases. Chapter 6 provides conclusions and suggestions for future work.

CHAPTER 2

LITERATURE REVIEW

Regression analysis is performed to learn the relationship between input and output variables. Input variables can be dependent on the output variable linearly or non-linearly. To establish the relationship between input and output variables for the linear dependency, linear regression (Seber and Lee, 2012; Montgomery et al., 2012) is performed. For the non-linear dependency, non-linear regression (Bates and Watts, 1988) is pursued. In the case of categorical outputs, classification analysis is performed in place of regression analysis to establish such a relationship. Years after years, numerous methods have been developed to predict outputs based on the input data either by classification or regression analysis. Accurate prediction becomes pronounced, especially in the field of machine learning (ML) (Michie et al., 1994). Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for both classification and regression problems. Such uses of SVM in both classification and regression analysis are respectively known as support vector classification (SVC) and support vector regression (SVR) (Vapnik, 2013; Vapnik et al., 1997; Drucker et al., 1997; Vapnik and Vapnik, 1998; Gunn, S.R., 1998; Schölkopf et al., 2002). Besides the field of machine learning, SVR is gaining its application in the field of statistics (Christmann et al., 2009; Hable, 2012), and data mining (Do and Poulet, 2005). This thesis uses SVR for prediction with interval data.

2.1 Classification Analysis Considering Uncertainty

In both SVC and SVR, SVM is first trained with the available dataset to obtain a trained model, which is then used to test a new dataset. However, if the data used for training are contaminated with various types of uncertainty, new challenges arise in prediction. Some researchers experiment with the effects of uncertainty on prediction using different loss functions. Hinge loss and pinball loss functions are two commonly used loss functions in SVM-based prediction. Hinge loss is related to the shortest distance between the classes while the pinball loss function is related to the quantile distance. The hinge loss function is traditionally used in SVM classifier. However, Huang et al. (2013) used pinball loss function in the classification problem and defined it as pin-SVM. They acutely investigated the properties of the pinball loss function in

pin-SVM and showed that SVM classifier with the pinball loss function is preferable for its noise-insensitivity and robustness property compared to the hinge loss function. The pinball loss function is also introduced in Twin support vector machine (TSVM) classification problem (Khemchandani and Chandra, 2007) by Xu et al. (2016) instead of the hinge loss function. Use of the pinball loss function in TSVM is defined as Pin-TSVM by Xu et al. (2016). Before Xu et al. (2016), Peng (2011) proposed twin parametric-margin support vector machines (TPMSVM) which are capable of handling only heteroscedastic noises. However, Xu et al. (2016) showed that with the increase of noisy samples, the prediction accuracy of Pin-TSVM increases more compared to Pin-SVM and TPMSVM, which support the more robust property of Pin-TSVM.

Bi and Zhang (2005) proposed a general statistical framework to tackle the noise in input data. Their proposed framework is based on a probability modelling approach. Their proposed formulations are interpreted through simple geometric representation and solved efficiently with their developed algorithm. The algorithm is known as total support vector classification (TSVC) algorithm. Bi and Zhang (2005) finally compared the performance of the proposed method with standard SVM in the presence of noisy input data. Trafalis and Gilbert (2006) proposed a robust approach for both classification and regression problem to make the support vector machines unaffected by the potential noise in input data. They approached their robust formulations by merely adding a bounded perturbation in the form of a norm within the standard SVM framework. However, their analysis does not contain any probabilistic distribution assumption on the noise. Moreover, standard techniques of solving mathematical programming problem cannot be used to solve their proposed robust programming problem. Linear or second-order cone programming (SOCP) must be used to pursue their proposed robust problem depending on the linearly or nonlinearly separable cases, respectively. Carrizosa et al. (2007) also proposed a robust classification model that can deal with both the cases- noisy and interval-valued data - through separating hyperplanes. Their proposed model is a generalization of the robust formulations proposed by Trafalis and Gilbert (2006). Although Carrizosa et al. (2007) applied their proposed model in multi-class classification and binary classification problems, non-linear separability through kernels is yet to be examined in their proposed approach. In the linear, binary and robust classification approach proposed by El Ghaoui et al. (2003) for multi-dimensional intervals, hyperrectangles are used to bound the unknown data. This classifier is a conservative approach that minimizes the worst-case value of a given loss function.

Besides the approaches with support vector machines for classifying the noisy and interval data, approaches based on the neural network cannot be sidelined (Rossi and Conan-Guez, 2002). Interval data are also considered in discriminant analysis by Ishibuchi et al. (1990), and Silva and Brito (2006). Silva and Brito (2006) experimented with linear discriminant analysis in three different approaches. First one is a probabilistic approach that assumes uniform distribution in each of the observed intervals while the second one is with the interval descriptive vertices. The third one uses the midpoints of the intervals as well as the lengths of the corresponding intervals. Thus, the proposed approaches by Silva and Brito (2006) utilize the variable information inherent within the intervals in different ways.

2.2 Regression Analysis with Noisy Data

If we consider prediction through regression analysis in the presence of noisy and interval data, the literature is not as vast as that of classification. As mentioned earlier, Trafalis and Gilbert (2006) proposed robust regression approaches that use support vector machines and deal with noisy data. However, a robust approach with SVR for dealing with interval data is hardly present in literature. Tanaka and Lee (1998), Hwang et al. (2006), Hao (2009) proposed quadratic programming approach, support-vector-interval-regression machines (SVIRM), and v – support vector interval regression networks (v –SVIRN), respectively, to deal with noisy data in regression analysis. All these three approaches perform interval regression analysis where the test data is expected to lie within a predicted interval. In SVIRM, Hwang et al. (2006) used possibility theory in association with the principles of standard support vector regression. However, later, in his proposed v –SVIRN, Hao (2009) introduced a parameter v in SVIRN to make SVIRN insensitive to various loss functions. Both the accuracy and training time are better for v –SVIRN compared to robust SVIRM as shown by Hao (2009). Petit-Renaud and Denœux (2004) proposed a non-parametric regression analysis based on the Dempster-Shafer evidence theory (Gordon and Shortliffe, 1984) that consider data uncertainty and impreciseness. Their proposed non-parametric regression analysis is known as evidence regression (EVREG). In EVREG, interval regression analysis is performed utilizing the Pignistic probability (Pignistic probability is a probability that a rational person will assign to an option when required to make a decision) distribution function.

2.3 Regression Analysis with Interval Data

Cattaneo and Wiencierz (2011) first introduced a likelihood-based robust approach that utilizes the likelihood-based model introduced by Cattaneo (2007) and considers interval data for regression analysis. However, the robust approach of Cattaneo and Wiencierz (2011) is based on the traditional regression framework that minimizes the squared or absolute error in fitting the regression function. In their proposed likelihood-based robust approach, Cattaneo and Wiencierz (2011) deal with both the statistical uncertainty and indetermination through two parameters. These two parameters are respectively cutoff points of likelihood and probability of error in imprecise observations. Different values of these two parameters influence the spread of the sets of undominated regression functions (i.e., the imprecise result of the regression). In all, Cattaneo and Wiencierz (2011) mainly showed how different values of the two parameters result in different sets of undominated regression functions. Thus, a defined framework for prediction is missing in the likelihood-based robust regression method of Cattaneo and Wiencierz (2011). Later, Cattaneo and Wiencierz (2014) investigated the application of likelihood-based robust approach in linear regression in the presence of interval data.

Utkin and Coolen (2011) first proposed a method for regression with interval data that uses SVM. The generalized-SVR framework presented by Utkin and Coolen (2011) exploits the approach proposed by Petit-Renaud and Denœux (2004) for fuzzy belief assignment-based risk functionals. In connection with imprecise probability theory (Walley, 1991), the generalized-SVR framework also exploits the approach of Walter et al. (2007) for the sets of probability distributions of the random noises present in the interval data. However, Utkin and Coolen (2011) finally deduced a simple formulation that uses the standard-SVR framework; however, just with a clever manipulation in the risk functional. The manipulation involves minimizing or maximizing the risk functional through choosing the appropriate bound value (either lower or upper bound value) from each interval. However, manipulation only with the boundary points eventually fails to consider the various uncertainties inherently present in the interval data. Moreover, the framework is not suitable for presence of interval data in both the input and output. In other words, interval data must be present only in output or in input observations to apply the generalized-SVR framework proposed by Utkin and Coolen (2011). Later, Wiencierz and Cattaneo (2015) investigated the generalized-SVR approach for its validity in terms of Representer Theorem, which is followed by

the experimentation with the wine quality dataset. However, both papers (Utkin and Coolen, 2011; Wiencierz and Cattaneo, 2015) lack clear explication of schemes to choose either the minimax or minimin approach for better prediction accuracy. Thus, the presence of very few literature in SVR with interval data, and their little concern for prediction accuracy provide the impetus to pursue this research. Hence, this thesis is delved into developing different approaches for regression analysis with interval data, which use the SVR framework. Also, the prediction accuracies and computational expenses of all the proposed approaches are compared. In the next chapter, basic concepts used in our proposed approaches are discussed in detail.

CHAPTER 3

BACKGROUND CONCEPTS REVIEW

3.1 Support Vector Regression

Support vector regression (SVR) and simple regression model have similitude in the sense that both minimize the regression error in defining the regression function; however, they differ in the techniques of minimizing the regression error. SVR tries to fit all the training data within a certain threshold and compensate for the data not fitted in the considered threshold. At the same time, SVR also minimizes model complexity through a regularizer, which is denoted by \mathbf{w} . Eq. (3.1) represents the linear SVR model

$$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b, \text{ where } \mathbf{w}, \mathbf{x} \in \mathcal{H}, b \in \mathbb{R}, \quad (3.1)$$

where b is a constant offset and the training dataset \mathbf{X} are considered to be independently and identically distributed (iid). These training data can be defined as follows:

$$\mathbf{X} := \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots \dots \dots, (\mathbf{x}_n, y_n)\} \in \mathcal{H} \times \mathbb{R} \quad (3.2)$$

In Eqs. (3.1) and (3.2), \mathcal{H} is a dot product space (also known as Hilbert Space) where the (mapped) input patterns live (Schölkopf et al., 2002). In order to penalize for the non-fitted data during training, the risk functional $R[f]$ introduced in the framework of statistical learning theory is defined as follows:

$$R[f] = \int c(\mathbf{x}, y, f(\mathbf{x})) dP(\mathbf{x}, y) \quad (3.3)$$

In Eq. (3.3), $P(\mathbf{x}, y)$ is the probability measure which is assumed to be responsible for generating the observations in Eq. (3.2), $f(\mathbf{x})$ is the regression estimate, and c is the cost or loss function. However, as P is not known priori, considering that probability of occurrence of all the data is the same, empirical risk functional can be defined as shown in Eq. (3.4).

$$R_{\text{emp}}[f] = \frac{1}{n} \sum_{i=1}^n c(\mathbf{x}_i, y_i, f(\mathbf{x}_i)) \quad (3.4)$$

As mentioned earlier, to find the best regression line, SVR requires to satisfy the two criteria – fitting the training data as much as possible and minimizing \mathbf{w} . However, these two criteria are conflicting. The objective function with these two conflicting criteria is shown in Eq. (3.5), which is known as regularized risk functional. To get the best regression line, regularized risk functional is minimized during SVM training, which is comprised of the regularization term and the risk functional as shown in Eq. (3.5)

$$R_{\text{reg}}[f, \mathbf{w}] = \frac{1}{2} \|\mathbf{w}\|^2 + \mathcal{C} R_{\text{emp}}[f] \quad (3.5)$$

Substituting Eq. (3.4) into Eq. (3.5), the regularized risk functional now becomes

$$R_{\text{reg}}[f, \mathbf{w}] = \frac{1}{2} \|\mathbf{w}\|^2 + \mathcal{C} \frac{1}{n} \sum_{i=1}^n c(\mathbf{x}_i, y_i, f(\mathbf{x}_i)) \quad (3.6)$$

In Eq. (3.6), \mathcal{C} is a constant which trades off between marginalizing error and model complexity. In more in-depth insight, the smaller the regularization term, the flatter the hyperplane that eventually induces more errors. Thus, the flatness of the regularization term enlarges the value of risk functional. The opposite scenario emerges with a larger regularization term. Therefore, there should be a controller for these two opposite scenarios, which induces moderate complexity in the hyperplane and at the same time account for the errors due to the flatness of the hyperplane. \mathcal{C} performs the function of that controller. Assuming $\frac{\mathcal{C}}{n} = C$ in Eq. (3.6), we get

$$R_{\text{reg}}[f, \mathbf{w}] = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n c(\mathbf{x}_i, y_i, f(\mathbf{x}_i)) \text{ where } C = \frac{\mathcal{C}}{n} \quad (3.7)$$

Now, cost function $c(\mathbf{x}_i, y_i, f(\mathbf{x}_i))$ used so far can be defined in general as in Eq. (3.8) where ε is known as soft margin loss.

$$c(\mathbf{x}, y, f(\mathbf{x})) = \begin{cases} 0 & \text{for } |y - f(\mathbf{x})| \leq \varepsilon \\ \tilde{c}(|y - f(\mathbf{x})| - \varepsilon) & \text{otherwise} \end{cases} \quad (3.8)$$

Thus, c becomes the loss function \tilde{c} for $|y - f(\mathbf{x})| > \varepsilon$ as can be observed in Eq. (3.8). Assuming $y - f(\mathbf{x}) = \xi \in \mathbb{R}$, Eq. (3.8) becomes

$$c(\xi) = \begin{cases} 0 & \text{for } |\xi| \leq \varepsilon \\ \tilde{c}(|\xi| - \varepsilon) & \text{otherwise} \end{cases} \quad (3.9)$$

In the maximum likelihood sense, cost function $c(\mathbf{x}_i, y_i, f(\mathbf{x}_i))$ can be written as

$$c(\mathbf{x}, y, f(\mathbf{x})) = -\log p(y - f(\mathbf{x})) \quad (3.10)$$

After applying Eq. (3.10) for all the iid training data, the probability density model of the loss function values generally becomes

$$p(\mathbf{X}_f | \mathbf{X}) = e^{-\sum_{i=1}^n c(\mathbf{x}_i, y_i, f(\mathbf{x}_i))} \quad (3.11)$$

where p represents the density model and $\mathbf{X}_f := \{(\mathbf{x}_1, f(\mathbf{x}_1)), (\mathbf{x}_2, f(\mathbf{x}_2)), \dots, (\mathbf{x}_n, f(\mathbf{x}_n))\}$

The commonly used loss functions and the corresponding density models derived through Eq. (3.11) are shown in Table 3.1

Table 3.1 Common loss functions and corresponding density models

	Loss function	Density Model
ε - intensive	$c(\xi) = \xi _\varepsilon$	$p(\xi) = \frac{1}{2(1 + \varepsilon)} \exp(- \xi _\varepsilon)$
Laplacian	$c(\xi) = \xi $	$p(\xi) = \frac{1}{2} \exp(- \xi)$
Gaussian	$c(\xi) = \frac{1}{2} \xi^2$	$p(\xi) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \xi^2\right)$
Huber's robust loss	$c(\xi) = \begin{cases} \frac{1}{2\sigma} (\xi)^2 & \text{if } \xi \leq \sigma \\ \xi - \frac{\sigma}{2} & \text{otherwise} \end{cases}$	$p(\xi) \propto \begin{cases} \exp\left(-\frac{1}{2\sigma} \xi^2\right) & \text{if } \xi \leq \sigma \\ \exp\left(\frac{\sigma}{2} - \xi \right) & \text{otherwise} \end{cases}$
Polynomial	$c(\xi) = \frac{1}{p} \xi ^p$	$p(\xi) = \frac{p}{2 \Gamma(1/p)} \exp(- \xi ^p)$

Piecewise polynomial	$c(\xi) = \begin{cases} \frac{1}{p\sigma^{p-1}} (\xi)^p & \text{if } \xi \leq \sigma \\ \xi - \sigma \frac{p-1}{p} & \text{otherwise} \end{cases}$	$p(\xi) \propto \begin{cases} \exp\left(-\frac{\xi^p}{p\sigma^{p-1}}\right) & \text{if } \xi \leq \sigma \\ \exp\left(\sigma \frac{p-1}{p} - \xi \right) & \text{otherwise} \end{cases}$
-------------------------	--	---

Considering the loss functions shown in Table 3.1, Eq. (3.9) can also be presented as in Eq. (3.12), and we stick to this representation throughout the remainder of this thesis.

$$c(\xi) = \max\{0, \tilde{c}(|\xi| - \varepsilon)\} \quad (3.12)$$

Using the loss function defined in Eq. (3.12), regularized risk functional in Eq. (3.7) can be rewritten as

$$R_{reg}[f, \mathbf{w}] = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n c(\xi) \quad (3.13)$$

However, in this thesis, we only consider the ε – insensitive loss function, for which Eq. (3.12) is replaced by Eq. (3.14). Note that we can pursue our study with any other loss function. However, our focus in this work is not to experiment with different loss functions rather proposing an approach to pursue SVR with interval data. The ε – insensitive loss function is considered as this is one of the mostly used loss functions while approaching SVR; hence, we stick to this loss function for the sake of illustration only. If we look for graphical representation of the ε – insensitive loss function, we get a ε – tube as shown in Figure 3.1, where the training data lying outside the ε – tube generate errors.

$$c(\xi) = |\xi|_{\varepsilon} = \max\{0, (|\xi| - \varepsilon)\} \quad (3.14)$$

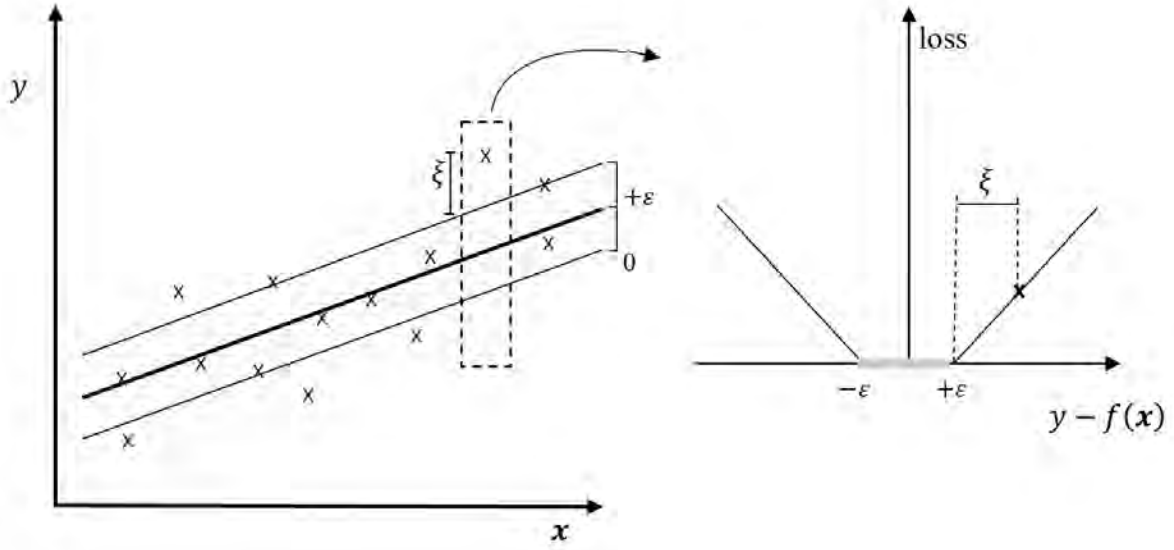


Figure 3.1 Soft margin loss setting for linear SVM in regression

The unconstrained optimization problem for minimizing the regularized risk functional (Eq. (3.13)) is usually transformed to the constrained optimization problem (Eq. (3.15)) for exploiting the advantage of quadratic programming problem. Such an advantage can be attained through the dual formulation considering that the problem in Eq. (3.15) is in the primal form. In Eq. (3.15), the cost functions are presented for the two cases - (i) predicted value is more than the actual value, (ii) predicted value is less than the actual value. In Eq. (3.15), $\xi_i^{(*)}$ stands for both ξ_i and ξ_i^* , which are limited to non-negative values only. If for each observation, different cost functions are chosen separately for ξ_i and ξ_i^* , representation of \tilde{c} become \tilde{c}_i and \tilde{c}_i^* respectively. However, this case is skipped in this thesis since ε – insensitive loss function is considered as the cost function.

$$\begin{aligned}
 \min_{\mathbf{w}, \xi^{(*)}, b} \quad & R_{emp}[f, \mathbf{w}] = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\tilde{c}(\xi_i) + \tilde{c}(\xi_i^*)) \\
 \text{s.t.} \quad & ((\mathbf{w}, \mathbf{x}_i) + b) - y_i \leq \varepsilon + \xi_i \\
 & y_i - ((\mathbf{w}, \mathbf{x}_i) + b) \leq \varepsilon + \xi_i^* \\
 & \xi^{(*)} \geq 0
 \end{aligned} \tag{3.15}$$

For the ε – insensitive loss function, the constrained optimization formulation in Eq. (3.15) becomes:

$$\begin{aligned}
\min_{\mathbf{w}, \xi^{(*)}, b} \quad & R_{emp}[f, \mathbf{w}] = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\
s.t. \quad & (\langle \mathbf{w}, \mathbf{x}_i \rangle + b) - y_i \leq \varepsilon + \xi_i \\
& y_i - (\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \leq \varepsilon + \xi_i^* \\
& \xi^{(*)} \geq 0
\end{aligned} \tag{3.16}$$

In order to formulate the dual of the primal form in Eq. (3.16), the Langrangian of Eq. (3.16) is taken as shown in Eq. (3.17). Lagrangian multipliers $\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \geq 0$ are introduced for the constraints of Eq. (3.16) as can be seen in Eq. (3.17).

$$\begin{aligned}
L = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) - \sum_{i=1}^n \eta_i \xi_i + \eta_i^* \xi_i^* \\
- \sum_{i=1}^n \alpha_i (\varepsilon + \xi_i + y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - b) \\
- \sum_{i=1}^n \alpha_i^* (\varepsilon + \xi_i^* - y_i + \langle \mathbf{w}, \mathbf{x}_i \rangle + b)
\end{aligned} \tag{3.17}$$

The Langrangian function in Eq. (3.17) has saddle point solutions with respect to both primal and dual variables. Hence, for duality, the gradient condition under KKT saddle point solutions of Eq. (3.17) becomes

$$\partial_b L = \sum_{i=1}^m (\alpha_i - \alpha_i^*) = 0 \tag{3.18}$$

Gradient condition:

$$\partial_{\mathbf{w}} L = \mathbf{w} - \sum_{i=1}^n (\alpha_i^* - \alpha_i) \mathbf{x}_i = 0 \tag{3.19}$$

$$\partial_{\xi^{(*)}} L = C - \alpha_i^{(*)} - \eta_i^{(*)} = 0 \tag{3.20}$$

Substituting Eq. (3.18) to Eq. (3.20) into Eq. (3.17) we get the dual form of Eq. (3.16) as shown in Eq. (3.21).

$$\begin{aligned}
\max_{\alpha_i^{(*)}} \quad & -\frac{1}{2} \sum_{i,j=1}^m (\alpha_i^* - \alpha_i) (\alpha_i - \alpha_i^*) \langle \mathbf{x}_i, \mathbf{x}_j \rangle - \varepsilon \sum_{i=1}^m (\alpha_i^* + \alpha_i) + \sum_{i=1}^m y_i (\alpha_i^* - \alpha_i) \\
s.t. \quad & \sum_{i=1}^m (\alpha_i^* - \alpha_i) = 0 \\
& \alpha_i^{(*)} \in [0, C]
\end{aligned} \tag{3.21}$$

The solution of Eq. (3.21) does not give us the value of bias or offset b . Determination of b requires calling the KKT orthogonality condition (Eq. (3.22)) at the optimal solution.

$$\begin{aligned}
& \alpha_i (\varepsilon + \xi_i + y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - b) = 0 \\
\text{Orthogonality} \quad & \alpha_i^* (\varepsilon + \xi_i^* - y_i + \langle \mathbf{w}, \mathbf{x}_i \rangle + b) = 0 \\
\text{Condition:} & \\
& \text{and}
\end{aligned} \tag{3.22}$$

$$\begin{aligned}
& (C - \alpha_i) \xi_i = 0 \\
& (C - \alpha_i^*) \xi_i^* = 0
\end{aligned} \tag{3.23}$$

The solution of Eq. (3.21) in conjunction with Eqs. (3.22) and (3.23) give us a clear idea of support vectors (SVs) as well as the offset value b . Performance of SVR relies on these support vectors (SVs) because the prediction accuracy of SVR does not depend on all the training data rather a part of the training data, which are known as SVs. Which training data act as support vectors (SVs) can be explained from the dual solution viewpoint. Three cases arise regarding the values of $\alpha_i^{(*)}$ – the decision variables of dual formulation in Eq. (3.21):

- (i) $\alpha_i^{(*)} = C$: In this case, left part of Eq. (3.23) becomes zero; therefore, $\xi^{(*)} \neq 0$. Hence, according to the definition of the ε – insensitive loss function, training data corresponding to this solution remain outside of the ε – tube (Figure 3.1)
- (ii) $\alpha_i^{(*)} = (0, C)$: In this case, as the first portion of Eq. (3.23) does not become zero, $\xi^{(*)}$ must be zeros. With $\xi^{(*)} = 0$, the right portion of Eq. (3.22) becomes $\varepsilon = \pm(y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - b)$, which indicate that training data now lie just on the boundary line of the ε – tube (Figure 3.1). Actually, from this case we can determine the value of the bias b as shown in Eq. (3.24).

$$b = y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - \varepsilon \text{ for } \alpha_i \in (0, C) \quad (3.24)$$

$$b = y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle + \varepsilon \text{ for } \alpha_i^* \in (0, C)$$

- (iii) $\alpha_i^{(*)} = 0$: This case results in $\alpha_i \alpha_i^* = 0$, which means that α_i and α_i^* cannot be nonzero at a time. If α_i, α_i^* vanish, second parts of Eq. (3.22) are not equal to zero, which implies $|y - f(\mathbf{x})| < 0$. In other words, in this case, the training data lie within the ε -tube (Figure 3.1) and result in zero losses. Hence, the training data lying within the ε -tube are kept out of consideration during the optimization. This conclusion also implicitly says that only for the case (i) and case (ii), training data contribute to the fitting of the hyperplane during solving the optimization problem in Eq. (3.21). Therefore, training data lying outside or on the boundary line of ε -tube are defined as SVs.

The lower the number of the training data participate in determining the hyperplane, the sparser the optimal solution, which eventually leads to faster training. Accordingly, the representation of the regularization term \mathbf{w} becomes sparse as it does not require all the training data to be considered. Such sparse representation of \mathbf{w} as shown in Eq. (3.25) is known as the so-called support vector expansion, i.e., \mathbf{w} can be completely described as a linear combination of the training patterns \mathbf{x}_i . In Eq. (3.25), \mathbf{w} is expressed as a linear combination of support vectors with coefficients $(\alpha_i^* - \alpha_i)$, which is actually the statement of the Representer theorem (Kimeldorf and Wahba, 1971).

$$\mathbf{w} = \sum_{i=1}^m (\alpha_i^* - \alpha_i) \mathbf{x}_i \quad (3.25)$$

Putting the expression of \mathbf{w} into Eq. (3.1), we get the expression of linear SVR model in support vector expansion form as shown in Eq. (3.26). Optimal values of $f(\mathbf{x})$ in Eq. (3.26) defines the so-called linear support vector machine (SVM).

$$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b = \sum_{i=1}^m (\alpha_i^* - \alpha_i) \langle \mathbf{x}_i, \mathbf{x} \rangle + b \quad (3.26)$$

The equations developed so far are for the linearly separable input space. However, in the presence of nonlinearly separable input space, one only needs to introduce a suitable feature space. This feature space maps the input space in such a way that in the feature space the mapped inputs are now easily separable by a hyperplane. Rewriting the dual formulation in Eq. (3.21), support vector expansion in Eq. (3.25), and linear SVR model in Eq. (3.26) in terms of mapped input space $\Phi(\mathbf{x})$, Eqs. (3.27), (3.28), and (3.29), respectively, can be obtained. Dot product of such mapped input space yields kernels, which are described in detail in the next section.

$$\begin{aligned}
\max_{\alpha_i^{(*)}} & -\frac{1}{2} \sum_{i,j=1}^m (\alpha_i^* - \alpha_i)(\alpha_i - \alpha_i^*) \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle - \varepsilon \sum_{i=1}^m (\alpha_i^* + \alpha_i) + \\
& \sum_{i=1}^m y_i (\alpha_i^* - \alpha_i) \\
\text{s.t.} & \sum_{i=1}^m (\alpha_i^* - \alpha_i) = 0 \\
& \alpha_i^{(*)} \in [0, C]
\end{aligned} \tag{3.27}$$

$$\begin{aligned}
\text{Support Vector Expansion} & \\
\text{using mapped input space} & \quad \mathbf{w} = \sum_{i=1}^m (\alpha_i^* - \alpha_i) \Phi(\mathbf{x}_i) \tag{3.28}
\end{aligned}$$

$$\begin{aligned}
\text{SVR Model using mapped} & \\
\text{input space} & \quad f(\mathbf{x}) = \sum_{i=1}^m (\alpha_i^* - \alpha_i) \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}) \rangle + b \tag{3.29}
\end{aligned}$$

3.2 Kernels

Kernel is mainly a functional measure of similarity or dissimilarity on the inputs, which can be obtained by taking dot product in the feature space. However, such measuring by a kernel is not restricted to SVM only. Certain classes of kernels induce feature spaces, which implies that certain feature space can be built based on the given kernel by exploiting the properties of that kernel function. The classes of kernels k considered in this thesis correspond to dot products in feature spaces \mathcal{H} via a map Φ where

$$\begin{aligned}
\Phi: \mathcal{X} &\rightarrow \mathcal{H} \\
\mathbf{x} &\mapsto \mathbf{x} := \Phi(\mathbf{x})
\end{aligned} \tag{3.30}$$

Hence,

$$k(\mathbf{x}, \mathbf{x}') = \langle \Phi(\mathbf{x}), \Phi(\mathbf{x}') \rangle \quad (3.31)$$

In Eq. (3.30), \mathcal{X} represents input domain, feature spaces \mathcal{H} is known as Hilbert Space. \mathcal{H} can be called Reproducing Kernel Hilbert Space (RKHS) when (i) for a given class of kernel, the related feature space can be generated, and (ii) the dot product of the feature space can reproduce the kernel itself. Reproducing the kernel becomes possible if the function $k: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ exists with the two properties - the reproducing property and the closed space property. These two properties are recited next from Schölkopf et al. (2002).

$$\langle f, k(\mathbf{x}, \cdot) \rangle = f(\mathbf{x}) \quad \forall f \in \mathcal{H} \quad (3.32)$$

Reproducing
property

In particular,

$$\langle k(\mathbf{x}, \cdot), k(\mathbf{x}', \cdot) \rangle = k(\mathbf{x}, \mathbf{x}') \quad (3.33)$$

Closed Space
property

\mathcal{H} is spanned by k , i.e., $\mathcal{H} = \overline{\text{span}\{k(\mathbf{x}, \cdot) | \mathbf{x} \in \mathcal{X}\}}$ where \bar{X} denotes the completion of the set X

In Eq. (3.32), f is a continuous evaluation functional on \mathcal{X} under Hilbert space. To obtain a vector space through mapping by f , f can be expressed in the form of linear combination as shown in Eq. (3.34)

$$f = f(\cdot) = \sum_{i=1}^n \gamma_i k(\cdot, \mathbf{x}_i) \quad (3.34)$$

In Eq. (3.34), the choice of coefficient γ is arbitrary. It follows directly from Eq. (3.33) that $k(\mathbf{x}, \mathbf{x}')$ is symmetric in its arguments and the satisfy the condition of positive definiteness. Here, note that a kernel matrix K (also known as Gram matrix) with its elements $K_{ij} := k(\mathbf{x}_i, \mathbf{x}_j)$ is not positive definite unless it has nonnegative eigenvalues in the presence of the symmetric property. Some of the most used kernel functions are polynomial kernels, Gaussian or radial basis kernels, and sigmoid kernel functions. They are defined as follows:

Polynomial Kernel: $k(\mathbf{x}, \mathbf{x}') = \langle \mathbf{x}, \mathbf{x}' \rangle^d$ where $d \in \mathbb{N}$ (3.35)

Gaussian or RBF kernel: $k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2\sigma^2}\right)$ where $\sigma \in \mathbb{R}$ (3.36)

Sigmoid kernel: $k(\mathbf{x}, \mathbf{x}') = \tanh(\kappa \langle \mathbf{x}, \mathbf{x}' \rangle + \theta)$ where $\kappa, \theta \in \mathbb{R}$ (3.37)

To show how kernel can be constructed from the vector space - which is generated with the images of the input pattern under mapping Φ , let's take the case of polynomial kernel considering $d = 2$ in Eq. (3.35)

$$\Phi: ([x]_1, [x]_2) \mapsto ([x]_1^2, [x]_2^2, [x]_1[x]_2, [x]_2[x]_1) \quad (3.38)$$

$$\begin{aligned} \langle \Phi(\mathbf{x}), \Phi(\mathbf{x}') \rangle &= [x]_1^2 [x']_1^2 + [x]_2^2 [x']_2^2 + 2[x]_1 [x]_2 [x']_1 [x']_2 \\ &= \langle \mathbf{x}, \mathbf{x}' \rangle^2 = k(\mathbf{x}, \mathbf{x}') \end{aligned} \quad (3.39)$$

Eq. (3.38) says that polynomial kernel is simply the square of the dot product in the input space. Similarly, for other types of kernel classes, one can build the desired kernel from the corresponding mapping with the help of Representer theorem and the kernel properties (Schölkopf et al., 2002). Dual formulation (Eq. (3.27)), support vector expansion (Eq. (3.28)) and the SVR model (Eq. (3.29)) are presented again respectively in Eqs. (3.40), (3.41) and (3.42) utilizing the kernel properties described in this section.

$$\begin{aligned} \max_{\alpha_i^{(*)}} \quad & -\frac{1}{2} \sum_{i,j=1}^m (\alpha_i^* - \alpha_i) (\alpha_i - \alpha_i^*) k(\mathbf{x}_i, \mathbf{x}_j) - \varepsilon \sum_{i=1}^m (\alpha_i^* + \alpha_i) + \\ & \sum_{i=1}^m y_i (\alpha_i^* - \alpha_i) \\ \text{s.t.} \quad & \sum_{i=1}^m (\alpha_i^* - \alpha_i) = 0 \\ & \alpha_i^{(*)} \in [0, C] \end{aligned} \quad (3.40)$$

Support Vector Expansion using Kernel $\mathbf{w} = \sum_{i=1}^m (\alpha_i^* - \alpha_i) k(\cdot, \mathbf{x}_i)$ (3.41)

SVR Model using Kernel $f(\mathbf{x}) = \sum_{i=1}^m (\alpha_i^* - \alpha_i) k(\mathbf{x}_i, \mathbf{x}) + b$ (3.42)

3.3 Nonlinear SVR in Primal Form

The best regression function through SVR is mostly searched in the dual form; however, the primal form can also be pursued for the same. Two major reasons can be mentioned for solving the problem in the dual: (i) the duality theory provides a convenient way to deal with the

constraints; (ii) the dual optimization problem can be written in terms of dot products, thereby, making it possible to use kernel function. Chapelle (2007) showed that these two causes can be accounted as the advantages of solving the dual problem but not as the limitation of solving the primal problem. With the help of Representer theorem, they also showed that solving the primal problem could be as efficient as the dual one for both linear and nonlinear cases. For an efficient solution of the primal form, an efficient algorithm is proposed by Chapelle (2007). This efficient algorithm requires the constrained form of primal problem transformed in the unconstrained form. The unconstrained form of the primal problem is already demonstrated in Eq. (3.7) of Section 3.1. Using f in place of \mathbf{w} , Eq. (3.7) can be redefined in terms of the objective function of the unconstrained primal optimization problem as shown in Eq. (3.43).

$$\min_{f, b} \frac{1}{2} \|f\|^2 + C \sum_{i=1}^n c(\mathbf{x}_i, y_i, f(\mathbf{x}_i)) \quad (3.43)$$

In Eq. (3.43), $f(\mathbf{x}_i) = \langle f, k(\mathbf{x}_i, \cdot) \rangle_{\mathcal{H}} + b$. Now, taking the derivative of the Lagrangian of Eq. (3.43) with respect to f and equating it to zero we get,

$$f^* + C \sum_{i=1}^n \frac{\partial c}{\partial f}(\mathbf{x}_i, y_i, f^*(\mathbf{x}_i)) k(\mathbf{x}_i, \cdot) = 0 \quad (3.44)$$

$$f^* = \sum_{i=1}^n \beta_i k(\mathbf{x}_i, \cdot) \quad (3.45)$$

where $\beta_i = -C \sum_{i=1}^n \frac{\partial c}{\partial \lambda}(\mathbf{x}_i, y_i, f(\mathbf{x}_i))$. Notice that, there is no difference in Eq. (3.45) with Eq. (3.41): β_i is just used in place of γ_i . However, β_i in Eq. (3.45) should not be interpreted as the Lagrange multipliers in Eq. (3.41). With the help of Eq. (3.45), the SVR model can be written as

$$f(\mathbf{x}) = \sum_{i=1}^n \beta_i k(\mathbf{x}_i, \mathbf{x}) + b \quad (3.46)$$

Using the definition of f from Eq. (3.45), unconstrained primal optimization problem in Eq. (3.43) becomes

$$\min_{\beta_i, \beta_j, b} \frac{1}{2} \sum_{i,j=1}^n \beta_i \beta_j k(\mathbf{x}_i, \mathbf{x}_j) + C \sum_{i=1}^n c \left(\mathbf{x}_i, y_i, \sum_{j=1}^n \beta_j k(\mathbf{x}_i, \mathbf{x}_j) + b \right) \quad (3.47)$$

For the ε –insensitive loss function, Eq. (3.47) takes the form

$$\min_{\beta_i, \beta_j, b} \frac{1}{2} \sum_{i,j=1}^n \beta_i \beta_j k(x_i, x_j) + C \sum_{i=1}^n \max \left\{ 0, \left| y_i - \left(\sum_{j=1}^n \beta_j k(x_i, x_j) + b \right) \right| - \varepsilon \right\} \quad (3.48)$$

That is,

$$\min_{\beta, b} \frac{1}{2} \beta^T K \beta + C \sum_{i=1}^n \max \{ 0, |y_i - (K_i^T \beta + b)| - \varepsilon \} \quad (3.49)$$

where K is a kernel matrix with its elements $K_{ij} = k(x_i, x_j)$ and K_i represents all the elements of column i . Eq. (3.49) can be easily solved by the updated rules of Newton optimization (Chapelle, 2007) or any similar or more efficient algorithm. For tuning the hyperparameters simultaneously with the optimization, conjoint optimization can be applied (Chapelle, 2007).

3.4 Probabilistic approach for dealing with interval data

To properly apprehend the uncertainty inherent in interval data, one should have a clear conception of uncertainty sources, and the resulting classification of uncertainty. Two sources of uncertainty are aleatory (irreducible) and epistemic (reducible) uncertainty. Aleatory uncertainty is the outturn of the natural reasons, and thereby irreducible. Examples of aleatory uncertainty include geometric tolerances, operating conditions, inherent randomness in material property, inherent variations in any physical process, etc. Noises in data actually represent the aleatory uncertainty. On the other hand, epistemic uncertainty arises mainly due to lack of knowledge, limited data or subjective data. History of research considering epistemic uncertainty is not so long. Even in the last decade, to many, the presence of epistemic uncertainty represented by intervals was not so common. For example, Ferson et al. (2004) in “Summary from the epistemic uncertainty workshop: consensus amid diversity” mentioned that before the workshop, many authors did not believe of any real situation where the information on a parameter lies in an interval. Later, Ferson et al. (2007) in detail scripted several situations where interval data may arise. The sources of epistemic uncertainty can be attributed to either stochastic but poorly known quantity or deterministic but poorly known quantity. Such uncertainty due to stochastic or deterministic but poorly known physical quantity is called statistical and subjective forms of epistemic uncertainty, respectively. The latter is also known as indetermination. In the former case, there is uncertainty about the distribution type as well as distribution parameters of the random

variable. Moment-based approach proposed in this paper focuses on the epistemic uncertainty about a stochastic but poorly known quantity.

Epistemic uncertainty arising from the interval data can be available both as a single interval and multiple intervals. However, as each interval is considered as an observation, dealing with multiple intervals is unnecessary for SVR with interval data. Note that unlike point data, moments of interval data are only available as bounds. Zaman et al. (2011) proposed moment-bounding algorithms to calculate the bounds on the first four moments of both single and multiple interval data. The bounds of the first four moments for single interval data are tabularized in Table 3.2 as developed by Zaman et al. (2011). Choosing a specific value arbitrarily from each of the intervals of the four moments results in a set of values of four moments and thus, such sets of moments can be infinite in number. Henceforth, we may fit infinitely many possible probability distributions to interval data. Again, as the distribution of the random variables described by interval data is also uncertain, the forcible use of a specific distribution for all sets of moments induces errors in the enumeration. Therefore, a flexible family of distributions should be used to fit interval data using moment matching approach. Pearson, Beta, Lamda, Johnson, etc. are some of the feasible four-parameter flexible family of distributions. We have used Johnson family of distributions in this thesis as it is a convenient choice for its easier transformation to a standard Gaussian space compared to others. Such an easier transformation to standard normal space is carried by Eq. (3.50)

$$z = \gamma + \delta f\left(\frac{x - \xi}{\lambda}\right) \quad (3.50)$$

where z stands for the standard normal variable; x is the variable in the original space; $\xi, \lambda, \delta, \gamma$ are the four parameters of Johnson family of distributions. If $\frac{x - \xi}{\lambda} = y$, then $f(y)$ may have any of the forms shown in Eq. (3.51).

$$f(y) = \begin{cases} \ln(y + \sqrt{y^2 + 1}) & \text{for unbounded } (S_U) \\ \ln\left(\frac{y}{1-y}\right) & \text{for bounded } (S_B) \\ \ln(y) & \text{for lognormal } (S_L) \\ y & \text{for normal } (S_N) \end{cases} \quad (3.51)$$

However, distribution to be chosen from the Johnson family of distributions depends on a set of point values of moments. An appropriate Johnson distribution is then identified with the help of Figure 3.2. Such identification requires the value of $\beta_1 \equiv m_3^2/m_2^3$, and $\beta_2 \equiv m_4/m_2^2$ where m_2, m_3, m_4 represent the second, third, and fourth moments, respectively.

Table 3.2 Methods for calculating moment bounds for single interval data

Moment	Condition		Formula
	Lower bound	Upper bound	
1	PMF = 1 at lower endpoint = 0 elsewhere	PMF = 1 at upper endpoint = 0 elsewhere	$M_1 = E(x)$
2	PMF = 1 at any point = 0 elsewhere	PMF = 0.5 at each endpoint	$M_2 = E(x^2) - (E(x))^2$
3	PMF = 0.2113 at lower endpoint = 0.7887 at upper endpoint	PMF = 0.7887 at lower endpoint = 0.2113 at upper endpoint	$M_3 = E(x^3) - 3E(x^2)E(x) + 2(E(x))^3$
4	PMF = 1 at any point = 0 elsewhere	PMF = 0.7887 at one of the endpoints = 0.2113 at other endpoints	$M_4 = E(x^4) - 4E(x^3)E(x) + 6E(x^2)(E(x))^2 - 3(E(x))^4$

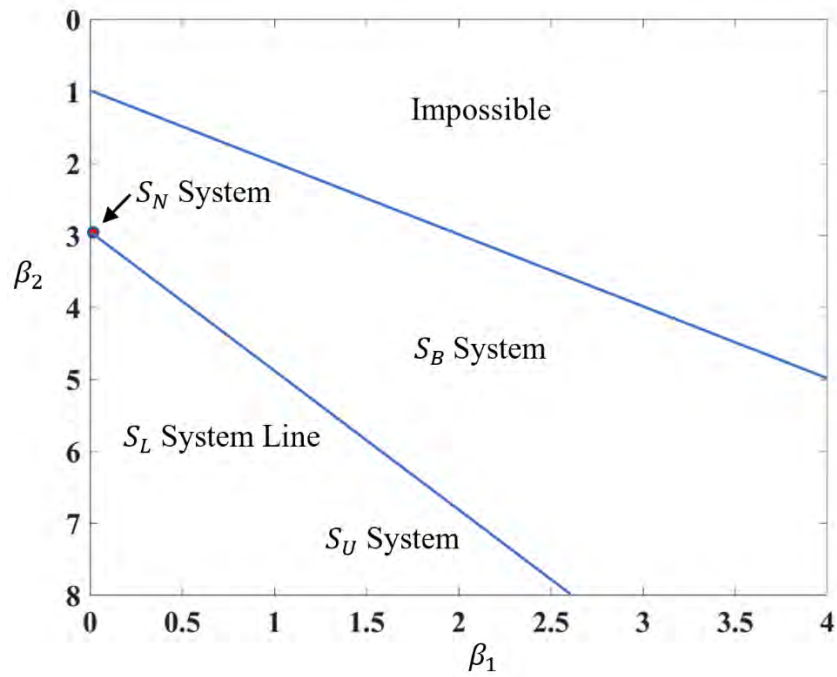


Figure 3.2 Identification of Johnson distribution family

In the next chapter, the four approaches proposed in this thesis to pursue SVR with the interval data are illustrated in detail. All the basic concepts introduced in this chapter are interchangeably exploited for exploring the proposed approaches in the next chapter.

CHAPTER 4

PROPOSED METHODOLOGIES

Before delving into the details of our proposed approaches, diagnosing how interval data may appear in the training data during regression is quite necessary. Based on the sources mentioned in Section 3.4, it is comprehensible that such uncertainty may arise in both input and output observations. Therefore, we consider three cases based on the presence of the interval data. Three cases are as follows: *(i)* interval data are present in input observations only; *(ii)* interval data are present in output observations only; *(iii)* interval data are present in both input and output observations. Note that it sounds quite impractical that interval data are always present in all the observations for the considered three cases. Therefore, in our proposed methodologies, we also take care of the scenarios that arise due to presence of both interval and point-valued observations in the same explanatory or response variable. In particular, we propose four different approaches. These approaches are the *(i)* moment-based approach, *(ii)* equiprobability-based approach, *(iii)* boundary-points-based approach, and *(iv)* extended generalized-SVR method. First two approaches are apposite for all the considered cases. The third one is apt when there is presence of interval data in the output, i.e., for both the second and third cases. The last one is just an intelligible extension of the generalized-SVR method developed by Utkin and Coolen (2011) for the presence of interval data in both input and output observations. We describe the proposed approaches in detail with all the adjoining variations that may arise in all three different cases. Note that the main challenge in pursuing SVR with interval data is to find the competent point data from the observed intervals while keeping all the salient properties of standard SVR unperturbed. From this aspect, all the approaches proposed in this thesis concur. In other words, in all our proposed approaches, all the optimal point data undergo training with standard SVR approach once after those point data are found out from the corresponding intervals through different schemes.

4.1 Moment-based Approach

Manipulation pursued in this approach is only with the moments of the interval data, which is performed outside of Standard-SVR. Thus, the existence and uniqueness for strict convexity property of standard-SVR representation is preserved here. In this approach, a two-loop nested

optimization problem is pursued. The outer loop of the nested optimization problem finds the competitive point data with the help of the moments, which are next sent to standard-SVR in the inner loop. Note that minimization of regularized risk functional (Eq. (3.7)) always goes on in the inner loop. Therefore, the target of the outer loop is to find out those point data from the corresponding intervals that minimize or maximize the minimized regularized risk functional in the inner loop. The moment-based approach finds those competitive point data based on moments.

The moment-based approach starts with enumerating bounds of moments for each interval-valued observation. Moment bounds require calculating the bounds for all the first four moments - mean, variance, skewness and kurtosis. Hence, the number of moment bounds is four times the number of interval-valued observations. The bounds of the first four moments for the interval-valued observations are determined by the moment bounding methods developed in Zaman et al. (2011). To be precise, we only need the help of Table. 3.2 for enumerating the bounds of moments of the interval-valued observations. Once we have the bounds of four moments available for each interval-valued observation, the algorithm in the outer loop chooses specific values of the moments from the corresponding bounds. Algorithm suited to solve the optimization problem that contains non-linear objective function and the bounds of the decision variables as only constraints can be used in the outer loop. In this thesis, sequential quadratic programming (SQP) is used in the outer loop. After the algorithm in the outer loop chooses a set of the moment values for each interval-valued observation, a suitable distribution from the Johnson family of distributions is determined through the moment matching approach. Next, a sample is drawn from the fitted Johnson distribution using the parameters of the corresponding distribution. Determining the parameters of a particular distribution from the Johnson family of distributions requires solving an optimization problem. This optimization problem uses that particular set of moments which have been used for selecting the distribution from the Johnson family. As samples, first, the standard normal variables are generated which are then used to generate samples in the desired distribution space through Eqs. (3.50) and (3.51). If the coordinate with the values of β_1 , and β_2 falls in the impossible area of Figure 3.2 (Section 3.4), a high value of regularized risk functional is assigned. In short, in the moment-based approach, mathematically a two-loop nested optimization as shown in Eq. (4.1) is formulated. In Eq. (4.1), we can see that the outer optimization problem is in search of the competitive point data. These point data are chosen from the corresponding intervals of interval-valued explanatory, and/or response variables with the help

of the moments \mathbf{m}_Z . The inner optimization problem in Eq. (4.1) then conducts standard-SVR training with the point data of explanatory and/or response variables provided by the outer loop.

$$\begin{aligned}
& \min_{\mathbf{w}, \xi^{(*)}, b} \left(\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \right) \\
& \text{s.t.} \quad (\langle \mathbf{w}, \mathbf{x}_i \rangle + b) - y_i \leq \varepsilon + \xi_i \\
& \quad \quad y_i - (\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \leq \varepsilon + \xi_i^* \\
& \quad \quad \mathbf{Z}_l^{x_i} \leq \mathbf{m}_Z^{x_i} \leq \mathbf{Z}_u^{x_i} \\
& \quad \quad \mathbf{Z}_l^{y_i} \leq \mathbf{m}_Z^{y_i} \leq \mathbf{Z}_u^{y_i} \\
& \quad \quad \xi_i^{(*)} \geq 0
\end{aligned} \tag{4.1}$$

In Eq. (4.1), $\mathbf{m}_Z = [\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3, \mathbf{m}_4]^T$ represents the vector of four moments for each interval data with \mathbf{Z}_l and \mathbf{Z}_u being the corresponding lower and upper bounds of the moments. Superscript in \mathbf{m}_Z , \mathbf{Z}_l , and \mathbf{Z}_u are used to denote whether the moments and their corresponding bounds are for interval-valued explanatory variables or response variables. The pseudocode of the proposed moment-based approach is shown in Figure 4.1.

Given the input observations \mathbf{x} and output observations y for training

Determine the bounds of moments \mathbf{Z}_l and \mathbf{Z}_u for each interval-valued observation using Table 3.2

repeat

Choose the competitive moment values \mathbf{m}_z within their corresponding bounds (\mathbf{Z}_l and \mathbf{Z}_u) for each interval-valued observation

for $i = 1$ to n (*number of interval data*) **do**

Determine a suitable distribution from Johnson family of distributions using Figure 3.2

Determine the point value within the interval using Eq. (3.50)

end for

Train all the point data with the standard-SVR and obtain the regularized risk functional value

until convergence

return the optimal sets of moment values for the interval-valued observations

Determine the point values for all the interval-valued observations using the corresponding sets of optimal moment values.

Train a model with the standard-SVR using the point values of all the training data

return the trained model

Figure 4.1 Pseudocode for the moment-based approach

The moment-based approach is applicable for all the three cases that we deal with in this thesis. Note, however that, in practical scenarios interval data may be available as the ranges of

some numerical values (e.g., persons' income range), i.e., categorized data. In other words, observed interval data can be categorical type. If all the interval data present in the output observations are such categorical types, then it is of no value of pursuing regression instead of classification. However, the coexistence of point data with the interval-valued observations in response variable or any explanatory variable makes the situation different where regression has to be pursued. Now in Eq. (4.1), whether only minimization or maximization or both should be pursued requires in-detail examination. Here, note that we perform point prediction of the unobserved values when there is no interval-valued observation in output. On the other hand, we predict unobserved output in an interval form rather than its point value when interval data are present in the output observations. Therefore, prediction of the bounds of the interval, within which the unobserved output is expected to lie, requires both the minimization and maximization of Eq. (4.1). Minimization yields one of the bounds while maximization yields another bound of the predicted interval. However, when there is no interval-valued output observation, we only need to pursue either minimization or maximization in Eq. (4.1) to predict a point value of the unobserved output. If point prediction is performed for both maximization and minimization of Eq. (4.1) separately, it can be seen that minimization yields better prediction accuracy. This is because, in minimization, those explanatory variables are searched from their corresponding intervals for which the minimized regularized risk functional in the inner loop of Eq. (4.1) is further minimized. Conversely, maximization increases the value of the minimized regularized risk functional, which eventually is reflected in the prediction accuracy.

The computational expense of the moment-based approach is a grave issue. The computational expense of the moment-based approach becomes vast, especially when interval data are present in both input and output observations. In such case, the optimization problem to obtain the parameters of the selected Johnson distribution family must be approached for both the interval-valued input and output variables, which makes each iteration of the nested optimization problem of Eq. (4.1) obsequiously time-consuming. To reduce the time consumption, we design a strategical pathway which also acts as the main scheme in proposing extended generalized-SVR approach (see Section 4.4). The idea is simple: conduct the nested optimization in Eq. (4.1) separately for interval-valued input and output observations to lessen the time consumption in each iteration. This idea is defined as the separation strategy. Thus, it is required to pursue Eq. (4.1) twice to complete an iteration with a separation strategy. In other words, an iteration under

separation strategy is completed in two steps. In each step, Eq. (4.1) is pursued either with interval-valued input observations or output observations. In the first step of the first iteration, one needs to start searching the point values from interval-valued input observations through Eq. (4.1). For the interval-valued outputs, arbitrary point values from the corresponding intervals are used. In the next step, point values for the interval-valued output observations are searched through Eq. (4.1). Now, as the point values of the interval-valued input observations, optimal inputs obtained from the first step are used. Similarly, optimal outputs obtained from the second step of the first iteration are used as the point values for the first step of the second iteration and so on. Thus, iterations go on until the termination criteria are satisfied. As a termination criterion, we consider the optimized value of minimized risk functional. If the optimized value of the minimized risk functional approaches toward an unanticipated direction in any step of an iteration compared to its corresponding step of the previous iteration, we need to stop. By an unanticipated direction, we mean that the minimum/maximum of the minimized risk functional increases/decreases in the present step of an iteration compared to the corresponding step of the previous iteration. Interestingly, if approaching toward an unanticipated direction starts in any step of an iteration, the untowardness continues for all the next steps of all the iterations. Moreover, it can be empirically shown that only at the second step of the second iteration, approaching to the unanticipated direction begins. Therefore, continuing after the first step of the second iteration becomes unnecessary in the separation strategy. Thus, in the separation strategy, Eq. (4.1) is pursued multiple times. However, the separation strategy solves the problem of computational intractability of the moment-based approach that arises when each iteration is executed with the input and output interval data all at a time. The pseudocode for the separation strategy in the moment-based approach is shown in Figure 4.2.

Given the input observations \mathbf{x} and output observations y for training

Assign point values for the interval-valued outputs arbitrarily from their corresponding bounds y^l and y^u

Repeat (Optimization to choose point values for interval-valued inputs)

Determine the bounds of moments \mathbf{Z}_l^x and \mathbf{Z}_u^x for each interval-valued input observation using Table 3.2

repeat

Choose the competitive moment values \mathbf{m}_z^x within their corresponding bounds (\mathbf{Z}_l^x and \mathbf{Z}_u^x) for each interval-valued input observation

for $i = 1$ to m (*number of interval data in input observations*) **do**

Determine a suitable distribution from Johnson family of distributions using Figure 3.2

Determine the point value within the interval using Eq. (3.50)

end for

Train all the point data with the standard-SVR and obtain the regularized risk functional value

until convergence

return the optimal sets of moment values for the interval-valued input observations

Determine the point values for all the interval-valued input observations using the corresponding sets of optimal moment values and assign these point values for all the interval-valued input observations

Continued ...

Determine the bounds of moments \mathbf{Z}_l^y and \mathbf{Z}_u^y for each interval-valued output observation using Table 3.2

Repeat (Optimization to choose point values for interval-valued outputs)

Choose the competitive moment values \mathbf{m}_z^y within their corresponding bounds (\mathbf{Z}_l^y and \mathbf{Z}_u^y) for each interval-valued output observation

for $i = 1$ to n (number of interval data in output observations) **do**

Determine a suitable distribution from Johnson family of distributions using Figure 3.2

Determine the point value within the interval using Eq. (3.50)

end for

Train all the point data with the standard-SVR and obtain the regularized risk functional value

until convergence

return the optimal sets of moment values for the interval-valued output observations

Determine the point values for all the interval-valued output observations using the corresponding sets of optimal moment values and assign these point values for all the interval-valued output observations for next iteration.

until convergence

Train a model with the standard-SVR using the point values of all the training data

return the trained model

Figure 4.2 Pseudocode for the moment-based approach with separation strategy

Besides the huge computational expenses of the moment-based approach, another shortcoming is its inability to handle missing data. As no definite bound is present for the missing data, the bounds of the moments cannot be obtained under the moment-based approach for missing data case. Thus, the moment-based approach cannot be applied for the missing data. However, the moment-based approach can be pursued after dropping the missing data from the observations. Despite these limitations of the moment-based approach, we investigate it in detail for examining its prediction accuracy in comparison with the other approaches proposed in this thesis.

4.2 Equiprobability-based Approach

The equiprobability-based approach formulates a two-loop nested optimization problem like the moment-based approach. Like the moment-based approach, it also executes standard-SVR training in the inner loop. However, in the equiprobability-based approach, the outer loop chooses the point values from the respective interval-valued observations in a different manner. In the equiprobability-based approach, all the point data within an interval are assumed to have an equal probability of appearing. Hence, the outer loop chooses the point data from the respective intervals on an equiprobability basis. Once point data are chosen, they are passed into the inner loop to undergo training with standard SVR. Eq. (4.2) shows the mathematical formulation of the equiprobability-based approach where all the notations remain the same as introduced previously

$$\begin{aligned}
& \min_{\mathbf{x}, \mathbf{y}} / \max_{\mathbf{w}, \xi^{(*)}, b} \left(\min_{\mathbf{w}, \xi^{(*)}, b} \left(\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \right) \right) \\
& s.t. \quad (\langle \mathbf{w}, \mathbf{x}_i \rangle + b) - y_i \leq \varepsilon + \xi_i \\
& \quad y_i - (\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \leq \varepsilon + \xi_i^* \\
& \quad \mathbf{x}_i^l \leq \mathbf{x}_i \leq \mathbf{x}_i^u \\
& \quad y_i^l \leq y_i \leq y_i^u \\
& \quad \xi_i^{(*)} \geq 0
\end{aligned} \tag{4.2}$$

in Chapter 3. The pseudocode of the proposed equiprobability-based approach is shown in Figure 4.3.

Given the input observations \mathbf{x} and output observations \mathbf{y} for training

Assign the lower and upper bounds of the interval-valued input observations into \mathbf{x}^l and \mathbf{x}^u , respectively.

Assign the lower and upper bounds of the interval-valued output observations into \mathbf{y}^l and \mathbf{y}^u , respectively.

Repeat

 Choose competitive point data within their bounds on an equiprobability basis for all interval-valued observations

 Train all the point data with the standard-SVR and obtain the regularized risk functional value

until convergence

return the optimal point values for the interval-valued observations

Train a model with the standard-SVR using the point values of all the training data

Return the trained model

Figure 4.3 Pseudocode for the equiprobability-based approach

Note that equiprobability-based approach does not require to determine the point values for the interval-valued variables based on their moments like the moment-based approach. Hence, the equiprobability-based approach is computationally more efficient compared to the moment-based approach. In particular, time consumption in each iteration of equiprobability-based approach is much less compared to the moment-based approach. Hence, for interval-valued observations present in both inputs and outputs, computational burden appeared in the equiprobability-based approach is not that much pronounced. Therefore, the equiprobability-based approach can be easily pursued without separation strategy, unlike the moment-based approach. However, we pursue the equiprobability-based approach in both manners - with and without the separation

strategy. The former is performed for comparing the performance of equiprobability approach with that of moment-based approach while pursued with the separation strategy. Missing data in the observations can be handled by the equiprobability-based approach. It is done through drawing data arbitrarily from an infinite real-valued space for the missing values. However, in the equiprobability-based approach, the rules of exploiting minimization and/or maximization remain the same as prescribed for the moment-based approach in Section 4.1.

4.3 Boundary-point-based Approach

The boundary-point-based approach is applicable only when interval data are present in the output observations. Hence, application of the boundary-point-based approach is not as universal as the moment-based and equiprobability-based approaches. This approach works with the boundary points of the interval-valued output observations when interval data may be present only in responses or in both explanatory and response variables. For the presence of the interval data in response only, the boundary-point-based approach simply pursues training with standard-SVR for twice, one with the lower bounds and the other with the upper bounds of the interval-valued output observations. If explanatory variables are also present in the interval form, moment-based or equiprobability-based approach are pursued to choose the point values of those explanatory variables. Inner loops of Eqs. (4.3) and (4.4) represent standard SVR formulations under the boundary-point-based approach that use lower and upper bounds of interval-valued responses, respectively. Interval data present in the input observations along with the output observations are handled with the outer loops of Eqs. (4.3) and (4.4). The moment-based approach is used in the outer loop of Eqs. (4.3) and (4.4) for choosing point values of explanatory variables. However, if the equiprobability-based approach is used, Eqs. (4.3) and (4.4) look like Eqs. (4.5) and (4.6), respectively. In Eqs. (4.3) to (4.6), \underline{y}_i and \bar{y}_i stand for the lower and upper bound of each interval-valued output observation, respectively. All other notations used in Eqs. (4.3) to (4.6) bear the same representation as they did before in this thesis. Note that in the boundary point-based approach, it is not required to perform both maximization and minimization to predict the interval where the unobserved output is expected to lie. Only minimization is pursued for Eqs. (4.3) - (4.6).

The main difference in the boundary-point-based approach with the moment-based and equiprobability-based approaches is worth mentioning. In the boundary-point-based approach,

instead of searching for the competitive point data for the interval-valued output observations, again and again, we simply use the boundary points of the intervals in two separate optimizations. Thus, the boundary-point-based approach is more efficient compared to the moment-based and the equiprobability-based approach. Also, the separation strategy as undertaken in the moment-based approach and equiprobability-based approach is of no use here, which can be discerned from the formulations in Eqs. (4.3) - (4.6). However, like the moment-based approach, for any missing response data, the boundary-point-value-based methodology should be pursued after stripping those missing observations from the training data.

$$\begin{aligned}
& \min_{\mathbf{m}_Z^x} \left(\min_{\mathbf{w}, \xi^{(*)}, b} \left(\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \right) \right) \\
& \text{s.t.} \quad (\langle \mathbf{w}, \mathbf{x}_i \rangle + b) - \underline{y}_i \leq \varepsilon + \xi_i \\
& \quad \underline{y}_i - (\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \leq \varepsilon + \xi_i^* \\
& \quad \mathbf{z}_l^{x_i} \leq \mathbf{m}_Z^{x_i} \leq \mathbf{z}_u^{x_i} \\
& \quad \xi_i^{(*)} \geq 0
\end{aligned} \tag{4.3}$$

$$\begin{aligned}
& \min_{\mathbf{m}_Z^x} \left(\min_{\mathbf{w}, \xi^{(*)}, b} \left(\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \right) \right) \\
& \text{s.t.} \quad (\langle \mathbf{w}, \mathbf{x}_i \rangle + b) - \bar{y}_i \leq \varepsilon + \xi_i \\
& \quad \bar{y}_i - (\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \leq \varepsilon + \xi_i^* \\
& \quad \mathbf{z}_l^{x_i} \leq \mathbf{m}_Z^{x_i} \leq \mathbf{z}_u^{x_i} \\
& \quad \xi_i^{(*)} \geq 0
\end{aligned} \tag{4.4}$$

$$\begin{aligned}
& \min_{\mathbf{x}} \left(\min_{\mathbf{w}, \xi^{(*)}, b} \left(\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \right) \right) \\
& \text{s.t.} \quad (\langle \mathbf{w}, \mathbf{x}_i \rangle + b) - \underline{y}_i \leq \varepsilon + \xi_i \\
& \quad \underline{y}_i - (\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \leq \varepsilon + \xi_i^*
\end{aligned} \tag{4.5}$$

$$\begin{aligned}
& \mathbf{x}_i^l \leq \mathbf{x}_i \leq \mathbf{x}_i^u \\
& \xi_i^{(*)} \geq 0 \\
\min_{\mathbf{x}} & \left(\min_{\mathbf{w}, \xi^{(*)}, b} \left(\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \right) \right) \tag{4.6} \\
s.t. & \quad (\langle \mathbf{w}, \mathbf{x}_i \rangle + b) - \bar{y}_i \leq \varepsilon + \xi_i \\
& \quad \bar{y}_i - (\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \leq \varepsilon + \xi_i^* \\
& \quad \mathbf{x}_i^l \leq \mathbf{x}_i \leq \mathbf{x}_i^u \\
& \quad \xi_i^{(*)} \geq 0
\end{aligned}$$

Pseudocode for the boundary-point-based method is shown in Figure 4.4.

Given the input observations \mathbf{x} and output observations \mathbf{y} for training

if \mathbf{x} and \mathbf{y} both contain interval data **then**

Determine point values of interval-valued input observations either by moment-based or equiprobability based approach

if moment-based approach is chosen

Use Eq. (4.3) for training with Standard SVR using lower bound values of interval-valued outputs

Use Eq. (4.4) for training with Standard SVR using upper bound values of interval-valued outputs

else if equiprobability-based approach is chosen

Use Eq. (4.5) for training with Standard SVR using lower bound values of interval-valued outputs

Continued ...


```

    Use Eq. (4.6) for training with Standard SVR using upper bound values of interval-
    valued outputs

    end

else if only  $y$  contain interval data

    Train with standard-SVR using lower bound values of interval-valued outputs

    Train with standard-SVR using upper bound values of interval-valued outputs

end if

return both the trained models

```

Figure 4.4 Pseudocode of Boundary-point-based approach

In the boundary-point-based method, presence of point data along with the interval data in the response variable is a matter of concern. In such case, for presence of the interval data in both input and output observations, training should be pursued after being concerned about two scenarios: (i) all the interval-valued input observations have their response variables interval-valued and vice versa; (ii) some of the interval-valued explanatory variables have their responses as point-valued and vice versa. For the first scenario, only the interval-valued observations are used for pursuing Eqs. (4.3) and (4.4), or Eqs. (4.5) and (4.6). Once the optimal point values of explanatory variables become handy, all the observed data go through training with standard SVR. However, such an approach cannot be followed straightforwardly for the second scenario. In the second scenario, using the point data of response variables with the corresponding interval-valued explanatory variables directly into Eqs. (4.3) and (4.4), or Eqs. (4.5) and (4.6) may be an unwise decision. This is because such use may affect the decision hyperplane of SVM in such a way that ultimately affects the prediction accuracy negatively. The same difficulty may arise when there are interval-valued observations only in the output where also point data coexist. A simple strategy to handle such situation is to convert those point-valued responses into intervals. Such intervals can be constituted using average length of the existing intervals and considering the coexisting

point-valued observations as the center of the corresponding intervals. However, this strategy requires further experimentation.

4.4 Extended generalized-SVR approach

The generalized-SVR model developed by Utkin and Coolen (2011) can be applied for regression with interval-valued observations in the training data. For interval data are present only in the response variables, Utkin and Coolen (2011) proposed the minimin and minimax formulations as shown in Eqs. (4.7) and (4.8), respectively. As we see from Eq. 4.7 that bound values of interval-valued outputs (either lower bound or upper bound value for a constraint) are used in the constraints in such a way that eventually yields greater value of loss function. To be more specific, use of lower bound \underline{y}_i instead of upper bound \bar{y}_i in the first constraint of Eq. (4.7) yield greater value of ξ_i . Similarly, greater value of ξ_i^* is engendered by the use of upper bound value \bar{y}_i in the second constraint of Eq. (4.7). The same scheme is utilized for the minimin approach in Eq. (4.8) but with the concern of smaller values of loss functions. Both the formulations in Eqs. (4.7) and (4.8) are called generalized-SVR because they are the generalization of the standard-SVR approach. In other words, standard-SVR can be deduced from them easily by rewriting Eqs. (4.7) and (4.8) with $\underline{y}_i = \bar{y}_i = y_i$, which happens when there are no interval data in the observations.

$$\begin{array}{l}
 \text{Minimax} \\
 \text{approach}
 \end{array}
 \begin{array}{l}
 \min_{\mathbf{w}, \xi^{(*)}, b} \quad \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\
 \text{s.t.} \quad ((\mathbf{w}, k(\mathbf{x}_i, \cdot)) + b) - \underline{y}_i \leq \varepsilon + \xi_i \\
 \bar{y}_i - ((\mathbf{w}, k(\mathbf{x}_i, \cdot)) + b) \leq \varepsilon + \xi_i^* \\
 \xi^{(*)} \geq 0
 \end{array} \tag{4.7}$$

$$\begin{array}{l}
 \text{Minimin} \\
 \text{approach}
 \end{array}
 \begin{array}{l}
 \min_{\mathbf{w}, \xi^{(*)}, b} \quad \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\
 \text{s.t.} \quad ((\mathbf{w}, k(\mathbf{x}_i, \cdot)) + b) - \bar{y}_i \leq \varepsilon + \xi_i
 \end{array} \tag{4.8}$$

$$\begin{aligned} \underline{y}_i - (\langle \mathbf{w}, k(\mathbf{x}_i, \cdot) \rangle + b) &\leq \varepsilon + \xi_i^* \\ \xi^{(*)} &\geq 0 \end{aligned}$$

The approaches used by Utkin and Coolen (2011) for the presence of interval data in output observations can also be developed for interval data present in explanatory variables; however, they are not as straightforward as they are for the case of response variables. This is because of the probable various natures of the regression function $f(\mathbf{x})$. Hence, similarly developed constraints from Eqs. (4.7) or (4.8) for the case of interval-valued explanatory variables may not work in a similar manner. Still, the minimax approach can be devised in somewhat similar manner with a clever strategy. However, for the minimin approach, this strategy does not work at all. For interval data present in explanatory variables only, the minimax approach looks like Eq. (4.9). We can see that Eq. (4.9) has four constraints instead of the two constraints in Eq. (4.7). Thus, like the \bar{y}_i and \underline{y}_i on the two constraints of Eq. (4.7), $\bar{\mathbf{x}}_i$ and $\underline{\mathbf{x}}_i$ are not directly introduced in Eq. (4.9) with only two constraints. This is because such an introduction does not always confirm the maximum of risk functional for minimax approach. For the minimin approach with the interval-valued explanatory variables, a different strategy is pursued where an unconstrained optimization problem is dealt with. In this strategy, the risk functional part of the unconstrained optimization problem of Eq. (3.43) requires only to be properly manipulated. In such manipulations, the boundary value (either lower bound or upper bound) of each interval-valued observation that results in the minimum of the risk functional is used. The resultant unconstrained problem can then be easily solved in the primal form with the help of techniques discussed in Section 3.3.

$$\begin{aligned} \min_{\mathbf{w}, \xi^{(*)}, b} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\ \text{s.t.} \quad & (\langle \mathbf{w}, k(\bar{\mathbf{x}}_i, \cdot) \rangle + b) - y_i \leq \varepsilon + \xi_i \\ & (\langle \mathbf{w}, k(\underline{\mathbf{x}}_i, \cdot) \rangle + b) - y_i \leq \varepsilon + \xi_i \\ & y_i - (\langle \mathbf{w}, k(\bar{\mathbf{x}}_i, \cdot) \rangle + b) \leq \varepsilon + \xi_i^* \\ & y_i - (\langle \mathbf{w}, k(\underline{\mathbf{x}}_i, \cdot) \rangle + b) \leq \varepsilon + \xi_i^* \\ & \xi^{(*)} \geq 0 \end{aligned} \tag{4.9}$$

Minimax approach

The main advantage of the generalized-SVR method is its superb efficiency compared to the other two approaches proposed in Sections 4.1 and 4.2. However, one of the main limitations of this approach is its inapplicability for interval data present in both input and output observations. Although the accuracy of this approach is not that much notable, swiftness of the generalized-SVR framework is tempting. Hence, we propose the extended generalized-SVR framework suitable for the presence of interval data in both input and output observations. The main strategy behind the proposed extended generalized-SVR framework is that if somehow the interval-valued input and output observations can be dealt with separately, the generalized-SVR approach can be exploited. The same perception is already utilized for overcoming the sluggishness of the iteration in the moment-based approach when interval data are present in both input- and output observations. Thus, in the proposed extended generalized-SVR approach, we also utilize the separation strategy to deal with the interval-valued explanatory and response variables separately. Termination condition of the separation strategy used in moment-based approach is also applicable for the extended generalized-SVR. The pseudocode of the extended generalized-SVR method is shown in Figure 4.5. Here, it is mentionable that in this thesis all the minimin and maximax approaches under the generalized-SVR are solved in their corresponding constrained or unconstrained primal forms.

Given the input observations x and output observations y for training

Choose bound values arbitrarily as the point data for interval-valued input observations

repeat

Solve maximax and minimin approach in Eqs. (4.7) and (4.8), respectively, for interval-valued output observations

Determine the point data of interval-valued output observations

Solve maximax approach in Eq. (4.9) for interval-valued input observations

Determine the point data of interval-valued input observations

Continued ...

until convergence

return the point values for the interval-valued observations

Train with standard-SVR using the point values of all the training data

return the trained model

Figure 4.5 Pseudocode of extended generalized-SVR approach

In the next chapter, we examine all our methodologies proposed in this chapter through four well-known datasets. For the case of interval data present only in output observations, we present one real and one synthetic dataset while for the other cases, one real dataset is used for each.

CHAPTER 5

NUMERICAL EXPERIMENTATION

In this chapter, all the proposed methodologies described in the previous chapter are in detail investigated for their prediction accuracy and computational efficiency with well-known datasets. We approach the detailed investigation case-wise where three cases are stratified based on the presence of interval data as mentioned earlier in Chapter 4. These three cases are (i) interval data present in input observations only, (ii) interval data present in output observations only, (iii) interval data present in both input and output observations.

5.1 Case 1: Interval Data Present in Input Observations Only

For this case, we consider the concrete slump test dataset (Yeh, 2007). This dataset is freely available at the UC Irvine Machine Learning Repository. This dataset contains seven input variables and three output variables. However, we consider only three input variables and one output variable for our investigation. Input variables represent the amount of cement, slag and water in kilograms in per cubic meter of concrete while output stands for 28-day compressive strength of concrete in MPa. Although all the observations in this dataset are point-valued, we have converted the input point data into intervals. Such conversion does not sound quite illogical as the measurement device may have its uncertainty in measurement and in showing the corresponding reading. We obtain the lower bounds of the interval data by subtracting 5% values of the original point data from the corresponding point data. Similarly, for the upper bound of the interval data, 5% values of the original point data are added with the corresponding point data. The concrete slump test data contain 103 observations of which 82 observations are considered as training data and the rest others are considered for testing purpose. Then the moment-based and the equiprobability-based approach have been examined for their prediction accuracies with the dataset. Boundary-point-based approach is not applicable for this dataset as there are no interval data in outputs. The prediction accuracies of the proposed two approaches are also compared with the existing generalized-SVR approach. Formulations of the proposed moment-based and equiprobability-based approach for the present case are presented in Eqs. (5.1) and (5.2), respectively. Eqs. (5.1) and (5.2) are the specialized forms of Eqs. (4.1) and (4.2), respectively,

because Eqs. (5.1) and (5.2) are the representations for the interval data present in input observations only. Eq However, the dual representation of the standard-SVR is used in Eqs. (5.1) and (5.2) where representations of all the notations remain unchanged. The Gaussian kernel is used as the kernel function as it is found to best describe the concrete slump test data.

$$\min/\max_{\mathbf{m}_Z^x} \left(\max_{\alpha_i^{(*)}} \left(-\frac{1}{2} \sum_{i,j=1}^m (\alpha_i^* - \alpha_i)(\alpha_i - \alpha_i^*)k(\mathbf{x}_i, \mathbf{x}_j) - \varepsilon \sum_{i=1}^m (\alpha_i^* + \alpha_i) + \sum_{i=1}^m y_i(\alpha_i^* - \alpha_i) \right) \right) \quad (5.1)$$

$$\begin{aligned} s.t. \quad & \sum_{i=1}^m (\alpha_i^* - \alpha_i) = 0 \\ & \mathbf{z}_l^{x_i} \leq \mathbf{m}_Z^{x_i} \leq \mathbf{z}_u^{x_i} \\ & \alpha_i^{(*)} \in [0, C] \end{aligned}$$

$$\min/\max_{\mathbf{x}} \left(\max_{\alpha_i^{(*)}} \left(-\frac{1}{2} \sum_{i,j=1}^m (\alpha_i^* - \alpha_i)(\alpha_i - \alpha_i^*)k(\mathbf{x}_i, \mathbf{x}_j) - \varepsilon \sum_{i=1}^m (\alpha_i^* + \alpha_i) + \sum_{i=1}^m y_i(\alpha_i^* - \alpha_i) \right) \right) \quad (5.2)$$

$$\begin{aligned} s.t. \quad & \sum_{i=1}^m (\alpha_i^* - \alpha_i) = 0 \\ & \mathbf{x}_i^l \leq \mathbf{x}_i \leq \mathbf{x}_i^u \\ & \alpha_i^{(*)} \in [0, C] \end{aligned}$$

For checking the prediction accuracy, we run both the maximization and minimization in the outer loop of Eqs. (5.1) and (5.2). Results obtained from the maximization and minimization problems are tabulated in Table 5.1 where it is observed that minimization yields less prediction error as expected. However, to get insights into such findings mathematically, we need to recall the basics of regularized risk functional in standard SVR formulation (Eq. 3.16). In the minimization of the regularized risk functional, the penalties for the training data not fitted within ε -tube are also considered alongside the searching for hyperplane parameter. During the

minimization in the outer loop of Eq. (5.1) or Eq. (5.2), the outer optimization algorithm searches the competitive point data from the intervals. Those competitive point data are chosen that set the

Table 5.1 Prediction errors of moment-based approach and equiprobability-based approach during minimization and maximization in the outer loop

Method	Prediction Errors (Mean Absolute Deviation, MPa)	
	Minimization in the outer loop	Maximization in the outer loop
Moment-based approach	4.75677	5.85975
Equiprobability-based approach	5.38466	6.99928

parameters of the decision hyperplane in such a way that most of the output training data are fitted within the ε –tube. Such strategy is aligned with the principal of standard SVR itself. For maximization opposite scenario happens. In other words, the outer algorithm searches in such a way that most of the output data points lie outside of ε –tube during training. Thus, minimization in the outer loop always yields less errors compared to its counterpart for the presence of interval data in explanatory variables only.

Table 5.2 shows the mean absolute deviation (MAD) of prediction errors as well as bounds

Table 5.2 Prediction errors for the concrete slump dataset using different approaches

Method	Prediction Error (Mean Absolute Deviation)	Bounds of Prediction Error
Moment-based approach	4.75677	[0.39160 12.14011]
Equiprobability-based approach	5.38466	[0.25399 14.60649]
Generalized-SVR approach	7.11436	[0.50894 15.48894]

of prediction errors for various methods. Bounds of prediction errors are the indicator of the minimum and maximum prediction errors happened during testing. Note that, such bounds do not provide enough conclusive remarks regarding the performance of different methods in terms of their prediction accuracy. Even sometimes they can be misleading. Hence, throughout this thesis, we follow the usual practice and thereby consider only the MAD of prediction errors while comparing different methods in terms of their prediction accuracy. Observing the average prediction accuracy of various methods from Table 5.2, we can see that the prediction accuracy of the moment-based approach is moderately greater than the equiprobability-based approach. However, such greater prediction accuracy is obtained at the cost of many times greater computational expenses. Of course, both the proposed approaches are better than the existing generalized-SVR approach to a great extent in terms of prediction accuracy. However, generalized-SVR approach consumes a very little computational time for training and testing purpose. The computational time required in the training phase, testing phase and their summation are shown in Table 5.3 for the moment-based, equiprobability-based and generalized-SVR approach. From Table 5.3, it is evident that the existing generalized-SVR surpasses both the

Table 5.3 Computational time of different approaches for the concrete slump dataset

Method	Training time (sec)	Testing Time (sec)	Total computational time (training + testing time) (sec)
Moment-based approach	145813.53513	0.00107	145813.53620
Equiprobability-based approach	1010.59422	0.00315	1010.59737
Generalized-SVR approach	50.50522	0.09396	50.59918

moment-based and equiprobability-based approach to a great extent in terms of its computational efficiency. One observation for this particular example is mention-worthy: although every iteration of the moment-based approach takes a lot more time compared to the equiprobability-based approach, the latter one ends up with more than ten times more iterations compared to the former

one. Still, for this particular dataset, the overall time consumption of the moment-based approach is many times more compared to the equiprobability-based approach. Thus, the use of the moment-based approach is more justifiable in offline learning where data needs not to be trained again and again like the online learning of data. The dataset considered in this example contains only 103 observations, thus, is of small size. However, an increase in the size of datasets enhance the convergence time of Eq. (5.1)/Eq. (5.2) a lot. Even, it may sometimes happen that data size is more than the dimension an algorithm can handle. To circumvent such computational intractability problem, suitable algorithms should be used, or some of the existing algorithms can be modified accordingly. However, it is out of the scope of this thesis and thus, opens the floor of future research.

5.2 Case 2: Interval Data Present in Output Observations only

For this case, unreliable sensor example has been considered, which is adapted from Petit-Renaud and Dencœux (2004). In this example, input x is generated using the following Eq. (5.3).

$$x_i = 0.5(i - 1), \quad i = 1, 2, \dots, N \quad (5.3)$$

Measurement values y given by the sensor can be determined by Eq. (5.4).

$$y_i \sim p_i \mathcal{N}(z_i, \sigma_i) + (1 - p_i) \mathcal{U}_{[0, x_N]} \quad (5.4)$$

In Eq. (5.4), the true output $z_i = x_i \sin x_i$; $p_i \in \mathcal{U}_{[0,1]}$ is the probability that the sensor will be in good operating condition during i^{th} observation; $\mathcal{U}_{[r_1, r_2]}$ is the uniform distribution in the range r_1 to r_2 ; \mathcal{N} is the normal distribution with mean z_i and standard deviation $\sigma_i \in \mathcal{U}_{[0.2, 2.2]}$. Outputs from Eq. (5.4) are point valued due to the point valued input variables. However, randomness in inputs induces randomness in measurement value y . Exploiting the random values shown by y , intervals of y with reasonable interval length have been built up for each observation i . In other words, for precise input values, we have now output values y available in the interval form. In order to experiment with our proposed methodologies with the unreliable sensor data, we use 201 observations that have only one input variable. Among the observations, 161 data are used in training and the rest 40 data are used for testing the trained model.

For the present case, we have used the boundary-point-based approach besides moment-based and equiprobability-based approaches. Formulations of the moment-based, equiprobability-based and boundary-point-based approaches for the present case are shown in Eqs. (5.5), (5.6), and (5.7) - (5.8), respectively. In all these formulations, it is considered that standard-SVR is solved in its dual form with Gaussian kernel. In Eq. (5.7), standard-SVR is pursued with the lower bounds of the interval-valued response variables while the upper bounds are used in Eq. (5.8). As, in the present case, interval data are present in output observations, we predict an interval instead of a point value of an unobserved output as mentioned earlier. Hence, we need to pursue here both the minimization and maximization in the outer loop of both the Eqs. (5.5) and (5.6). The rationale behind the preference between minimization and maximization for prediction accuracy still remains the same as discussed in Section 5.1 - maximization always yields more errors compared to minimization. Nevertheless, both are dealt here as we need to build the prediction intervals within which unobserved outputs are expected to lie. However, the predicted values obtained from minimizing the outer loop may be greater or less than those obtained from maximizing the outer loop. Hence, we build the lower bounds of the prediction intervals using the minimum between the values predicted by the minimization and maximization. Similarly, we use the maximum values for upper bounds.

$$\min/\max_{\mathbf{m}_z^y} \left(\max_{\alpha_i^{(*)}} \left(-\frac{1}{2} \sum_{i,j=1}^m (\alpha_i^* - \alpha_i)(\alpha_i - \alpha_i^*)k(\mathbf{x}_i, \mathbf{x}_j) - \varepsilon \sum_{i=1}^m (\alpha_i^* + \alpha_i) + \sum_{i=1}^m y_i(\alpha_i^* - \alpha_i) \right) \right) \quad (5.5)$$

$$s.t. \quad \sum_{i=1}^m (\alpha_i^* - \alpha_i) = 0$$

$$\mathbf{z}_l^{y_i} \leq \mathbf{m}_z^{y_i} \leq \mathbf{z}_u^{y_i}$$

$$\alpha_i^{(*)} \in [0, C]$$

$$\min/\max_y \left(\max_{\alpha_i^{(*)}} \left(-\frac{1}{2} \sum_{i,j=1}^m (\alpha_i^* - \alpha_i)(\alpha_i - \alpha_i^*)k(\mathbf{x}_i, \mathbf{x}_j) - \varepsilon \sum_{i=1}^m (\alpha_i^* + \alpha_i) + \sum_{i=1}^m y_i(\alpha_i^* - \alpha_i) \right) \right) \quad (5.6)$$

$$s.t. \quad \sum_{i=1}^m (\alpha_i^* - \alpha_i) = 0$$

$$y_i^l \leq y_i \leq y_i^u$$

$$\alpha_i^{(*)} \in [0, C]$$

$$\max_{\alpha_i^{(*)}} -\frac{1}{2} \sum_{i,j=1}^m (\alpha_i^* - \alpha_i)(\alpha_i - \alpha_i^*)k(\mathbf{x}_i, \mathbf{x}_j) - \varepsilon \sum_{i=1}^m (\alpha_i^* + \alpha_i) + \sum_{i=1}^m \underline{y}_i(\alpha_i^* - \alpha_i) \quad (5.7)$$

$$s.t. \quad \sum_{i=1}^m (\alpha_i^* - \alpha_i) = 0$$

$$\alpha_i^{(*)} \in [0, C]$$

$$\max_{\alpha_i^{(*)}} -\frac{1}{2} \sum_{i,j=1}^m (\alpha_i^* - \alpha_i)(\alpha_i - \alpha_i^*)k(\mathbf{x}_i, \mathbf{x}_j) - \varepsilon \sum_{i=1}^m (\alpha_i^* + \alpha_i) + \sum_{i=1}^m \bar{y}_i(\alpha_i^* - \alpha_i) \quad (5.8)$$

$$s.t. \quad \sum_{i=1}^m (\alpha_i^* - \alpha_i) = 0$$

$$\alpha_i^{(*)} \in [0, C]$$

MAD of prediction errors and computational time incurred by the three proposed approaches as well as the existing generalized-SVR approach are shown in Table 5.4 and Table 5.5, respectively. Table 5.4 also refers the prediction error bounds incurred by those methods. Like the previous case, prediction error (MAD) of the moment-based approach is the lowest among all as can be observed in Table 5.4. However, its prediction accuracy is not so much pronounced now if we alongside consider its huge computational expense. Existing generalized-SVR method predicts intervals through both the minimin and minimax approach. Although the existing approach is computationally efficient compared to the moment-based and equiprobability-based

Table 5.4 Prediction errors for the unreliable sensor problem using different approaches

Method	Prediction Error (Mean Absolute Deviation)	Prediction Error Bounds
Moment-based approach	2.39618	[0 8.76444]
Equiprobability-based approach	2.49415	[0 10.13397]
Boundary-point-based approach	2.63911	[0 10.79160]
Generalized-SVR approach	4.91341	[1 15.38839]

Table 5.5 Computational time of different approaches for the unreliable sensor problem

Method	Training time (sec)	Testing Time (sec)	Total computational time (training + testing time) (sec)
Moment-based approach	377416.65493	0.01929	377416.67422
Equiprobability-based approach	7314.93623	0.02409	7314.96032
Boundary-point-based approach	46.25860	0.01088	46.26948
Generalized-SVR approach	210.51572	0.00227	210.51799

approach, errors incurred in this approach are significant compared to all the proposed approaches. The average prediction accuracy of the boundary-points-based approach, though lower compared to the moment-based and equiprobability-based approach, still, is so close to them as can be shown from Table 5.4. Moreover, such prediction accuracy of the boundary-point-based approach is attained with much greater computational efficiency compared to the moment-based and equiprobability-based approaches. However, to gain more assurance on the prediction accuracy of

boundary-point-based approach as well as to comprehend its functional principles graphically, another numerical example under the present case is presented next.

The dataset considered for further experimentation with the boundary-point-based approach is wine quality dataset. This dataset was used by Wiencierz and Cattaneo (2015) for investigating the performance of the generalized-SVR approach. The considered dataset was collected to study the quality of Vinho Verde wines from Portugal and was initially introduced by Cortez et al. (2009) in a venture of modelling wine preferences. The dataset is now freely available at the UC Irvine Machine Learning Repository. There are 11 input variables, e.g., fixed acidity, volatile acidity, citric acidity, pH etc. in the datasets which are based on physicochemical tests. The output variable is the quality of the wine that is based on sensory data. Like Wiencierz and Cattaneo (2015), we create intervals $[0, 0.5]$, $[0.5, 1.5]$, $[1.5, 2.5]$, $[2.5, 3.5]$, $[3.5, 4.5]$, $[4.5, 5.5]$, $[5.5, 6.5]$, $[6.5, 7.5]$, $[7.5, 8.5]$, $[8.5, 9.5]$, $[9.5, 10]$, respectively, for the discrete values - 0,1,2,...,9,10 of sensory data on wine quality. Thus, this example contains categorical variables in interval-valued output observations. 679 and 232 observations are considered from this wine quality dataset for training and testing purpose, respectively. MAD of prediction errors as well as minimum and maximum errors incurred by the proposed and existing approaches are shown in Table 5.6 where linear kernel is used for prediction.

Table 5.6 Prediction errors for the wine quality dataset using different approaches

Method	Prediction Error (Mean Absolute Deviation)	Bounds of Prediction Errors
Moment-based approach	0.4056	[0 2]
Equiprobability-based approach	0.4071	[0 2]
Boundary-point-based approach	0.4158	[0 2]
Generalized-SVR approach	0.4946	[0 2]

The computational expenses of all the proposed approaches and the existing generalized-SVR approach are shown in Table 5.7 in terms of computational time. Like the previous two

datasets, time consumed by moment-based approach is many times more than the second most time-consuming equiprobability-based approach. Unlike the unreliable sensor problem, existing generalized-SVR approach here consumes the least computational time as can be seen from Table 5.7.

Table 5.7 Computational time of different approaches for the wine quality dataset

Method	Training time (sec)	Testing Time (sec)	Total computational time (training + testing time) (sec)
Moment-based approach	259227.25711	0.00467	259227.26178
Equiprobability-based approach	8201.69410	0.00426	8201.69836
Boundary-point-based approach	130.28392	0.00346	130.28738
Generalized-SVR approach	115.950165	0.003155	115.95331

As observed in Table 5.6, the average prediction accuracy of the boundary-point-based approach is slightly deviant compared to the more accurate moment-based and equiprobability-based approach. The prediction error of the generalized-SVR framework is again the most like the unreliable sensor problem. If we need to contemplate on the rationales of prediction by the boundary-based-approach, the hyperplane of SVM in standard SVR training should be recalled. Such hyperplane is decided through the maximal fitting of the training point data. Thus, two SVM hyperplanes are generated in the boundary-point-based approach, each one through training with one bound of the interval data. For the wine quality problem with one input and output variable, we simply get a line as the decision hyperplane of SVM. Now, as we know that for the interval-valued output observations, probable point values of outputs lie within the intervals. Accordingly, it can be anticipated that the future observations may lie within the boundary created by two SVM hyperplanes in the boundary-point-based approach. In Figure 5.1, we see that the point-valued test data lie within the predicted boundary lines attained from the boundary-point-based approach.

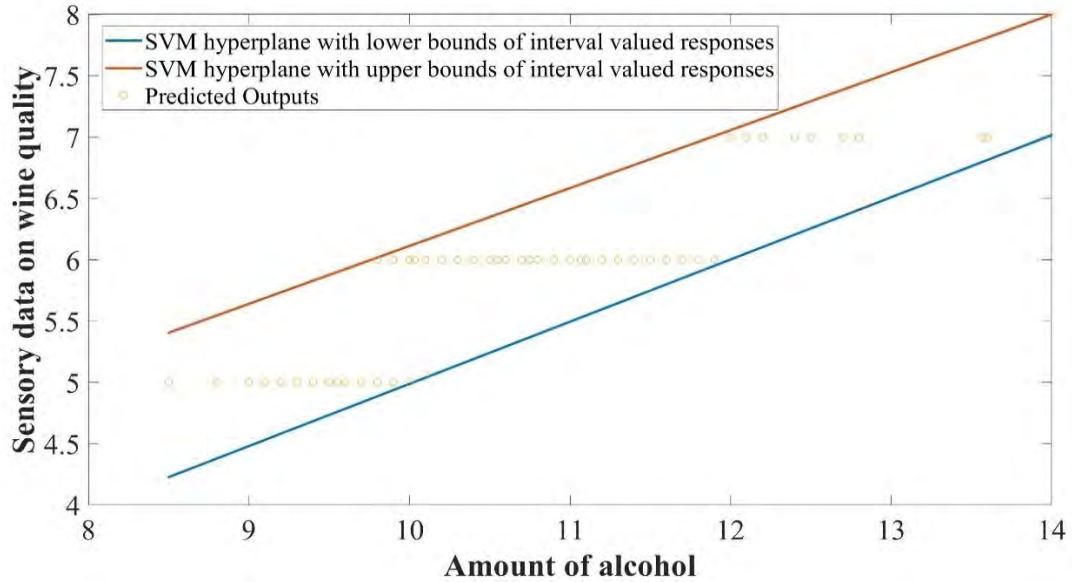


Figure 5.1 Prediction bounds from the boundary-point-based approach for the wine quality dataset

Note that the rationale just explored above becomes somewhat vague when the length of the interval is so wide. However, considering the sources of interval uncertainty in data science, it is not so much common to get such high levels of uncertainties that yield so wider length of intervals except for the category type intervals in some applications. Moreover, one may find some analogies between the boundary-point-based approach and the approach with the midpoints of the intervals and their lengths for training (Neto and de Carvalho, 2008). This issue is put ahead for extensive experimentation in future. Also, the issues regarding the applicability and the prediction accuracy of the boundary-point-based approach for the coexistence of point data with the interval-valued output observations require further experimentation as discussed in Section 4.3.

5.3 Case 3: Interval Data Present in Both Input and Output Observations

For this case, we collect data from “Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS)-German General Social Survey” of 2016, which is provided by GESIS-Leibniz Institute for the Social Sciences. This survey asked the participants many questions on many subjects starting from their age. For our experimentation, we choose only the surveyed person’s age, the number of months that surveyed person lived abroad, no of family members s/he has as the input variables and choose her/his per capita income in pounds (£) as the output variables

from the huge list of survey questionnaires. Here, data for months of living abroad are collected in categorical forms, e.g., up to 3 months, 4 to 6 months, and so on. Many of the surveyed persons preferred not to expose their income precisely rather gave answers in the income category provided in the survey questionnaires. Thus, we get interval data in output as well as in input. However, point data coexist with the interval-valued observations in the output. 214 and 42 observations are respectively considered for testing and training purpose from the social survey dataset. For this dataset, all the four proposed methodologies can be applied due to the presence of interval data in both input and output observations. In particular, besides the moment-based, equiprobability-based, boundary-point-based approach, we can also apply extended generalized-SVR for this dataset. Table 5.8 summarizes the MAD of errors as well as minimum and maximum errors incurred in predicting the intervals of per capita income for different proposed approaches when the linear kernel is considered to be best fitted for the social survey dataset. Note that, like the previous two cases, we only consider the MAD of prediction errors in Table 5.8 for comparison purpose.

Table 5.8 Prediction errors from all the proposed approaches for the social survey dataset

Method	Error (MAD) in per-capita income prediction	Prediction error bounds
Moment-based approach	27.82759	[0 51]
Equiprobability-based approach with the separation strategy	28.03448	[0 51]
Equiprobability-based approach without the separation strategy	31.20690	[0 51]
Boundary-points-based approach	9.51724	[0 25]
Extended generalized-SVR approach	40.06897	[0 70]

In the present case, if we consider the computational power available to us, pursuing the moment-based approach with both the interval-valued input and output observations at a time is very challenging. Hence, the moment-based approach is pursued using the separation strategy

which results in the formulations like Eqs. (5.1) and (5.5) for the interval-valued input and output observations, respectively. However, pursuing both interval-valued inputs and outputs at a time with the equiprobability-based approach is not a grave problem. Hence, we pursue it without the separation strategy. We also pursue the equiprobability-based approach with the separation strategy to compare its prediction accuracy with that of the moment-based approach. The formulations under equiprobability-based approach without the separation strategy look like Eq. (4.2) while the use of separation strategy arise when separation strategy is used for the prediction problem with social survey dataset yields Eqs. (5.2) & (5.6). More specifically, under the separation strategy, if outputs are considered as point-valued, Case-1 arises while Case-2 arises for the inputs considered as point-valued. Whether the separation strategy is exploited or not, minimization and maximization rules for all the formulations under the present case remain the same as they are for the first two cases in Section 5.1 and 5.2. Thus, whether it is moment-based approach or equiprobability-based approach, only minimization is used for interval-valued input observations. For the interval-valued output observations, both minimization and maximization are used. Note that as the interval data in the outputs are categorical type for the social survey dataset, one may only pursue minimization for interval-valued output observations. This is because, in the categorical type output, a particular category can be used as prediction when point-value predicted through minimization falls in that particular category. However, while pursuing only minimization, it should be concerned that some prediction accuracies may need to be compromised for not dealing the maximization problem. For example, when only minimization is used with the moment-based approach, prediction error becomes 33.06897. On the other hand, prediction error is only 27.82759 when both minimization and maximization is pursued under the moment-based approach. However, such increased prediction accuracy is sometimes attained through the so much enhanced length of the prediction intervals. Such wider length of prediction intervals may not be helpful to the decision-makers in many cases. Table 5.8 shows that with separation strategy, moment-based approach surpasses equiprobability-based approach in prediction accuracy marginally. However, equiprobability approach using all the interval-valued inputs and outputs at a time leads somewhat more prediction errors compared to when interval-valued inputs and outputs are dealt separately. The more prediction errors of the former one can be imputed to the participation of interval-valued inputs in the maximization process. As reasoning, it can be recalled that maximization with the interval-valued inputs increases the

prediction errors as explored in Section 5.1. Here, one may argue that that with the separation strategy prediction errors occurs due to the random choosing of point values from the corresponding intervals initially while it is not for the other one. However, such prediction error is surpassed by the advantages that the explanatory variables do not participate in the maximization process in the separation strategy and the eventual outcome is reflected in Table 5.8.

Exploiting the separation techniques for the generalized-SVR we can use generalized-SVR for the presence of interval data in both inputs and outputs. However, mentionable prediction error occurs here compared to other proposed approaches, which can be attributed mainly to the framework of generalized-SVR itself. As the generalized-SVR is pursued more than once here, prediction error increase. However, the computational efficiency of extended generalized-SVR approach in its prediction should never be overlooked. At this point, empirical findings regarding the performance (in terms of prediction accuracy) of the minimax and minimin approach under the generalized-SVR framework should be noted. Empirical findings say that for both the interval-valued input observations, minimax approach under generalized-SVR method works better while for the interval-valued output both the minimin and minimax approaches should be examined. Of course, for the non-categorical type of interval-valued output observations, both minimin and minimax approach need to be pursued to obtain the prediction interval like the unreliable sensor problem in Section 5.2.

It is surprising to observe the prediction accuracy of boundary-point-based approach. The large spread of the predicted intervals of the boundary-point-based approach can be attributed to its high prediction accuracy. Note that as we are predicting the intervals of unobserved output using the categorical type interval-valued observations, all our predictions also should be in that categorical form. This rule is followed in all the proposed approaches for this social survey dataset problem.

If we now focus on the computational efficiency of all the proposed approaches used for the social survey dataset, as intuitive, the least computational efficient approach is the moment-based approach. The computational time of all the proposed approaches can be observed in Table 5.9 where it shows that the equiprobability based approach is the second most time-consuming approach. Note that, in the enumeration of the computational time of the moment-based and

Table 5.9 Computational time of different approaches for the wine quality dataset

Method	Training time (sec)	Testing Time (sec)	Total computational time (training + testing time) (sec)
Moment-based approach	580800.548494	0.003285	580800.551779
Equiprobability-based approach	6852.767964	0.006793	6852.774757
Boundary-point-based approach	1526.017184	0.008398	1526.025582
Generalized-SVR approach	39.323944	0.017740	39.341684

equiprobability-based approaches, both the approaches are pursued with the separation strategy.

The next most time-consuming approach for this dataset is the boundary-point-based approach. However, difference between computational time consumed by the boundary point-based approach and extended generalized-SVR approach is now grave. This is because, here, alongside the training with the boundary points of the interval-valued outputs, moment-based or equiprobability-based approaches are also required for choosing the point values for the interval-valued inputs. However, in choosing such point data, it is wiser to use the computationally efficient equiprobability-based approach. Accordingly, equiprobability-based approach is used here within the boundary-point-based approach for the considered social survey dataset.

5.4 Discussion of Findings

The performances of the proposed approaches, in terms of prediction accuracy and computational efficiency, are evident from the findings shown in the last three sections. In general, we can say that moment-based approach performs better in terms of its prediction accuracy for most of the cases if we sideline the high prediction accuracy of the boundary-point-based approach obtained through a large prediction interval for the third case. However, question arises regarding

the application of the moment-based approach, since it is computationally expensive. However, extensive computational expense of this approach does not totally negate its application. Of course, as mentioned earlier, in online learning platform, where up-to-the-minute prediction is required and the prediction is performed in every hour iteratively, moment-based approach loses its usefulness. However, there are many instances where offline learning is the last resource for training and prediction. Aerospace design or any other constructional design are some of such instances. During designing a component, the decision-makers need to take decision with the data at hand. At the same time, they obviously anticipate the prediction approach that can help them make decision with the best accuracy possible. In such situation, the moment-based approach is obviously a great choice for prediction if there is presence of interval data. On the other hand, if accomplishing a prediction with the most recent data in a shorter time is the main concern, we need to choose an approach other than the moment-based approach. Equiprobability-based, boundary-point-based, extended generalized-SVR approach are the available options in this aspect. Use of equiprobability-based approach is discouraged in any emergency situation if the other two approaches can be used instead. This is because equiprobability-based approach consumes more time compared to the other two approaches for all the three cases presented in the last three sections (Sections 5.1 - 5.3). As an example of urgent situation, we can consider some unanticipated happenings during a flight of an aerospace vehicle when it is required to continuously screen the health of some crucial components of that vehicle. In such screening, prediction through online learning is an obvious requirement where boundary-point-based or extended generalized approach can play important roles. Thus, all our proposed approaches, though vary in their performances, have their varied applications according to the requirements demanded by different practical situations.

CHAPTER 6

CONCLUSIONS AND FUTURE SCOPES

6.1 Conclusions

The modern age is the era of data availability as well as the easiness of data collection in the online platform. However, contamination of the data by noise and impreciseness no way should be ignored in such an era. Mentionable sources of interval data are, but not limited to, data rounding, data binning, data heaping, data censoring, measurement instrument reading uncertainties, etc. Ignoring the uncertainty present in the interval data may become a grave issue for prediction accuracy, especially, in the field of statistical ML.

In this thesis, we have considered SVR for prediction and proposed four different approaches that deal with the interval uncertainty in prediction with SVR. The proposed approaches are moment-based approach, equiprobability-based approach, boundary-point-based approach, and extended generalized-SVR approach. However, they differ in their working principles, prediction accuracy, computational efficiency as well as applicability; all of these are in detail explored in this thesis. Boundary-point-based approach and extended generalized-SVR approach have their limitation regarding applicability. The former can be applied only for interval-valued responses while the later is specially developed to circumvent the limitations of the generalized-SVR approach in the presence of both interval-valued input and output observations. Both the moment-based and equiprobability-based approaches consider the statistical uncertainty in the interval data. They are also equally applicable for interval data present anywhere in input, or output, or both. However, the moment-based approach is observed to surpass the equiprobability-based approach in prediction accuracy somewhat but at the cost of huge numerical expenses. Concerns regarding the computational expenses of the proposed approaches are also addressed in terms of their computational time spent in training and testing. Four datasets – concrete slump dataset, unreliable sensor problem, wine quality dataset, social survey data from ALLUBUS-GESIS – are used for investigating the prediction accuracy and computational efficiency of the first three proposed approaches in terms of the minimization and maximization of the corresponding formulations. Moreover, the generalized-SVR framework is comprehended

in terms of the minimin and minimax approaches while pursuing its extension as the extended generalized-SVR. This extended generalized-SVR is examined with the social survey data for its prediction accuracy and computational efficiency also. In short, the boundary-point-based approach is always computationally more efficient compared to the moment-based and equiprobability-based approaches. For interval data present in both input and output observations, extended generalized-SVR outweigh other three proposed approaches in terms of its computational efficiency. On the other hand, the moment-based approach is always observed to outperform equiprobability-based and extended generalized-SVR approach in terms of prediction accuracy.

In all, the contributions of this thesis are broad. Addressing the vast gap in the literature of SVR with interval data, proposing four approaches with the focus on the prediction accuracy and computational expenses are the main contributions in short. All the concerns regarding the accuracy of the proposed approaches are explored in detail alongside measuring the computational time. All the three cases at which interval data may appear in practice are considered in detail with the well-known datasets while examining the proposed approaches. In addition, better prediction accuracy of all the proposed approaches compared to the existing approach highlights the importance of the study.

6.2 Future Scopes

Although the moment-based approach is somewhat better in its prediction accuracy in most of the cases, further studies are demanded for lessening its huge time consumption. The huge time consumption by the moment-based approach also limits its usage for offline training only. When it is the concern of tractability to deal with a huge dataset by the outer optimization algorithm of the proposed nested formulations, experimenting with different types of efficient algorithms can be recommended for the future. At this concern, various chunking type algorithms or algorithms that consider only two variables at a time (e.g., sequential minimal optimization (SMO)) can be designed accordingly. How the presence of missing values affects the prediction accuracies of the proposed approaches also require further experimentation. Note that, in the interval data, both the statistical and subjective uncertainty can be present. However, none of our proposed approaches deals with subjective uncertainty, i.e., indetermination. Thus, the possibility of incorporating the

indetermination in the proposed approaches and the consequent effects on the prediction accuracy can be investigated in future.

REFERENCES

- Basak, D., Pal, S. and Patranabis, D.C., 2007. Support vector regression. *Neural Information Processing-Letters and Reviews*, 11(10), pp.203-224.
- Bates, D.M. and Watts, D.G., 1988. *Nonlinear regression analysis and its applications* (Vol. 2). New York: Wiley.
- Bi, J. and Zhang, T., 2005. Support vector classification with input data uncertainty. In *Advances in neural information processing systems* (pp. 161-168).
- Carrizosa, E., Gordillo, J. and Plastria, F., 2007. Classification problems with imprecise data through separating hyperplanes. *MOSI Department, Vrije Universiteit Brussel, Tech. Rep. MOSI/33*.
- Cattaneo, M.E., 2007. Statistical decisions based directly on the likelihood function. *Doctoral dissertation, ETH Zurich*.
- Cattaneo, M.E. and Wiencierz, A., 2011. Robust regression with imprecise data. *Technical Report 114, Department of Statistics, LMU Munich*.
- Cattaneo, M.E. and Wiencierz, A., 2012. Likelihood-based imprecise regression. *International Journal of Approximate Reasoning*, 53(8), pp.1137-1154.
- Cattaneo, M.E. and Wiencierz, A., 2014. On the implementation of LIR: the case of simple linear regression with interval data. *Computational Statistics*, 29(3-4), pp.743-767.
- Chapelle, O., 2007. Training a support vector machine in the primal. *Neural computation*, 19(5), pp.1155-1178.
- Christmann, A., Steinwart, I. and van Messem, A., 2009. On consistency and robustness properties of support vector machines for heavy-tailed distributions. *Statistics and Its Interface*, 2(3), pp.311-327.
- Cortez, P., Cerdeira, A., Almeida, F., Matos, T. and Reis, J., 2009. Modeling wine preferences by data mining from physicochemical properties. *Decision Support Systems*, 47(4), pp.547-553.
- Do, T.N. and Poulet, F., 2005. Kernel methods and visualization for interval data mining. In *Proceedings of the Conference on Applied Stochastic Models and Data Analysis, ASMDA* (pp. 345-354).
- Drucker, H., Burges, C.J., Kaufman, L., Smola, A.J. and Vapnik, V., 1997. Support vector regression machines. In *Advances in neural information processing systems* (pp. 155-161).

- El Ghaoui, L., Lanckriet, G.R.G. and Natsoulis, G., 2003. Robust classification with interval data. *Technical Report UCB/CSD-03-1279, EECS Department, University of California, Berkeley.*
- Person, S., Joslyn, C.A., Helton, J.C., Oberkampf, W.L. and Sentz, K., 2004. Summary from the epistemic uncertainty workshop: consensus amid diversity. *Reliability Engineering & System Safety*, 85(1-3), pp.355-369.
- Person, S., Kreinovich, V., Hajagos, J., Oberkampf, W. and Ginzburg, L., 2007. Experimental uncertainty estimation and statistics for data having interval uncertainty. *Sandia National Laboratories, Report SAND2007-0939.*
- Gordon, J. and Shortliffe, E.H., 1984. The Dempster-Shafer theory of evidence. *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, 3, pp.832-838.
- Gunn, S.R., 1998. Support vector machines for classification and regression. *ISIS technical report*, 14(1), pp.5-16.
- Hable, R., 2012. Asymptotic normality of support vector machine variants and other regularized kernel methods. *Journal of Multivariate Analysis*, 106, pp.92-117.
- Hao, P.Y., 2009. Interval regression analysis using support vector networks. *Fuzzy sets and systems*, 160(17), pp.2466-2485.
- Heitjan, D.F. and Rubin, D.B., 1991. Ignorability and coarse data. *The annals of statistics*, pp.2244-2253.
- Hsu, C.W., Chang, C.C. and Lin, C.J., 2003. A practical guide to support vector classification.
- Hwang, C., Hong, D.H. and Seok, K.H., 2006. Support vector interval regression machine for crisp input and output data. *Fuzzy Sets and Systems*, 157(8), pp.1114-1125.
- Huang, X., Shi, L. and Suykens, J.A., 2013. Support vector machine classifier with pinball loss. *IEEE transactions on pattern analysis and machine intelligence*, 36(5), pp.984-997.
- Ishibuchi, h., Tanaka, h. and Fukuoka, n., 1990. Discriminant analysis of multi-dimensional interval data and its application to chemical sensing. *International Journal of General System*, 16(4), pp.311-329.
- Kimeldorf, G. and Wahba, G., 1971. Some results on Tchebycheffian spline functions. *Journal of mathematical analysis and applications*, 33(1), pp.82-95.
- Khemchandani, R. and Chandra, S., 2007. Twin support vector machines for pattern classification. *IEEE Transactions on pattern analysis and machine intelligence*, 29(5), pp.905-910.

- Michie, D., Spiegelhalter, D.J. and Taylor, C.C., 1994. Machine learning. *Neural and Statistical Classification*, 13.
- Montgomery, D.C., Peck, E.A. and Vining, G.G., 2012. *Introduction to linear regression analysis* (Vol. 821). John Wiley & Sons.
- Neto, E.D.A.L. and de Carvalho, F.D.A., 2008. Centre and range method for fitting a linear regression model to symbolic interval data. *Computational Statistics & Data Analysis*, 52(3), pp.1500-1515.
- Peng, X., 2011. TPMSVM: a novel twin parametric-margin support vector machine for pattern recognition. *Pattern Recognition*, 44(10-11), pp.2678-2692.
- Petit-Renaud, S. and Denœux, T., 2004. Nonparametric regression analysis of uncertain and imprecise data using belief functions. *International Journal of Approximate Reasoning*, 35(1), pp.1-28.
- Rossi, F. and Conan-Guez, B., 2002. Multi-layer perceptron on interval data. In *Classification, Clustering, and Data Analysis* (pp. 427-434). Springer, Berlin, Heidelberg.
- Schölkopf, B., Smola, A.J. and Bach, F., 2002. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press.
- Seber, G.A. and Lee, A.J., 2012. *Linear regression analysis* (Vol. 329). John Wiley & Sons.
- Smola, A.J. and Schölkopf, B., 2004. A tutorial on support vector regression. *Statistics and computing*, 14(3), pp.199-222.
- Silva, A.P.D. and Brito, P., 2006. Linear discriminant analysis for interval data. *Computational Statistics*, 21(2), pp.289-308.
- Tanaka, H. and Lee, H., 1998. Interval regression analysis by quadratic programming approach. *IEEE Transactions on Fuzzy Systems*, 6(4), pp.473-481.
- Trafalis, T.B. and Gilbert, R.C., 2006. Robust classification and regression using support vector machines. *European Journal of Operational Research*, 173(3), pp.893-909.
- Utkin, L.V. and Coolen, F.P., 2011, July. Interval-valued regression and classification models in the framework of machine learning. In *ISIPTA* (Vol. 11, pp. 371-380).
- Vapnik, V., Golowich, S.E. and Smola, A.J., 1997. Support vector method for function approximation, regression estimation and signal processing. In *Advances in neural information processing systems* (pp. 281-287).

- Vapnik, V. and Vapnik, V., 1998. Statistical learning theory Wiley. *New York*, pp.156-160.
- Vapnik, V., 2013. *The nature of statistical learning theory*. Springer science & business media.
- Walley, P., 1991. Statistical reasoning with imprecise probabilities. *Chapman and Hall*
- Walter, G., Augustin, T. and Peters, A., 2007, July. Linear regression analysis under sets of conjugate priors. In *ISIPTA* (Vol. 7, pp. 445-455).
- Wiencierz, A. and Cattaneo, M., 2015, July. On the validity of minimin and minimax methods for Support Vector Regression with interval data. In *9th international symposium on imprecise probability: Theories and applications* (pp. 325-332).
- Xu, Y., Yang, Z. and Pan, X., 2016. A novel twin support-vector machine with pinball loss. *IEEE transactions on neural networks and learning systems*, 28(2), pp.359-370.
- Yeh, I.C., 2007. Modeling slump flow of concrete using second-order regressions and artificial neural networks. *Cement and Concrete Composites*, 29(6), pp.474-480.
- Chapelle, O., 2007. Training a support vector machine in the primal. *Neural computation*, 19(5), pp.1155-1178.
- Zaman, K., Rangavajhala, S., McDonald, M.P. and Mahadevan, S., 2011. A probabilistic approach for representation of interval uncertainty. *Reliability Engineering & System Safety*, 96(1), pp.117-130.