



A new method of linear support vector regression with interval data

Mojtaba Baymani^{a,*}, Nima Salehi-M.^b, Hoda Saffaran^c

^aDepartment of Mathematics, Quchan University of Technology, Quchan, Iran

^bDepartment of Computer, Engineering Faculty, Ferdowsi University of Mashhad, Mashhad, Iran

^cDepartment of Mathematics, Islamic Azad University, Quchan Branch, Quchan, Iran

(Communicated by Saman Babaie-Kafaki)

Abstract

In this paper, the Linear support vector regression approach is proposed for solving the regression problem with interval data, which is called interval support vector regression (ISVR). The ISVR approach is equivalent to solving a linear constrained quadratic programming problem (QPP) with an interval cost coefficient in which the value of the objective function is in an interval. Instead of solving an interval QPP, we solve two QPPs and prove that the cost values of these two problems are the lower bound and the upper bound of the target value of the interval QPP. We show these two mentioned QPPs are equivalent to two support vector regression problems, in which the first problem uses from the lower bound of the data, and the second problem considers the upper bound of the data to obtain the regression function. Some experiments are made to demonstrate the performance of our method compared with the known algorithms on several artificial, benchmark and real practical datasets.

Keywords: Quadratic Programming, Computing methodologies and applications, Linear Regression.

2010 MSC: 90C20, 68Uxx, 62J05.

1. Introduction

Support vector machines (SVMs), as a new class of learning algorithms, have emerged for pattern classification and regression. The structural risk minimization principle is implemented in SVMs

*Corresponding author

Email addresses: m_baymani@qiet.ac.ir; mhbaymani@yahoo.com (Mojtaba Baymani), ni.salehi@mail.um.ac.ir (Nima Salehi-M.), h.saffaran@yahoo.com (Hoda Saffaran)

Received: December 2019 *Accepted:* January 2020

by minimizing the upper bound of the generalization error [5]. Support vector regression (SVR) has been applied to estimation data problems. SVR targets a linear regression through a data set, where points may be in feature space or in a higher space([1],[13]). The kernel trick is often used for obtaining nonlinear SVR, which maps the data to a higher dimensional space and applies a kernel function. But, in the higher space for large-scale data, the volume of computing increases sharply and the time-consuming has happened [6]. In the literature, the original form of SVR, namely ε -SVR, has been widely used. ε -SVR finds a function $f(x)$ such that more training samples set in the ε -intensive tube between $f(x) + \varepsilon$ and $f(x) - \varepsilon$.

Recently, some novel SVR algorithms have been proposed for data regression, consisting ε -twin SVR (ε -TSVR) [14], twin parametric insensitive SVR (TPISVR) ([12], [15]) and parametric-insensitive nonparallel SVR (PINSVR)[16]. These algorithms determine a pair of nonparallel proximal functions, $f_1(x)$ and $f_2(x)$, by solving two smaller sized QPPs instead of a function $f(x)$ solved of the larger QPP one in the ε -SVR, which make them have the faster learning speed than classical ε -SVR. For ε -TSVR, it determines two ε -insensitive proximal functions by using the ε -insensitive loss function, so it has better regression performance. These algorithms determine the regressor by averaging of $f_1(x)$ and $f_2(x)$. In the real world, many training data cannot be accurately determined

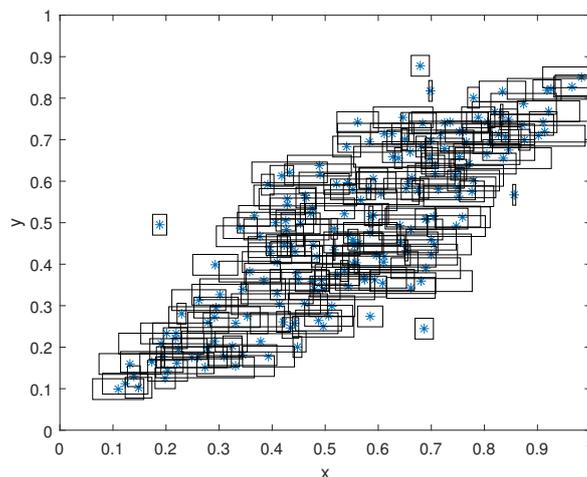


Figure 1: Sample of one-dimensional syntactic interval data points

due to the measurement error or the uncertainty of the information is existed. For noisy regression problems, training-error minimization may over train the noise into the learning model, thereby it leads to an unsatisfied test result. These problems motivate the use of SVR for the interval data that is presented in this paper. An example of the interval data is shown in Figure 1. The horizontal axis represents the interval data points and the vertical axis represents the label. ε -SVR and ε -TSVR both assume that the noise level of training data is uniform throughout the domain or some information about noise is given. In some problems that the noise is strongly dependent on the input, TPISVR and PINSVR obtain the parametric insensitive down- and up-bound functions by using the parametric-insensitive loss function. However, with newly added parameters, the size of problem is larger and so, the computation time is increased. Some literature has been considered a linear regression problem with interval-valued data. Billard et al. were one of the pioneers of research in this field [3]. They proposed the center method(CM) for obtaining a regression line by using data centers. Then, the regression function values in the lower bound and upper bound of a new interval data were considered as the estimation interval for observation [3]. Later, the researchers extended this method

and proposed the center and range methods (CRM). In these methods, two linear regressions are obtained by using the center and range of data(CRM). Then, by using them, the range of observation was estimated from a new interval value input ([9], [2], [10] and [8]). All of the mentioned methods were based on least squares estimation, and mathematically it cannot be assured that the lower bound is necessarily smaller than the upper bound. To overcome this problem, Neto et al. proposed a regression approach, named constraint center and range method (CCRM), by adding nonnegative constraints for coefficients in the CRM method[11]. In this paper, We extend the ε -SVR method for interval data and display it with ISVR. The ISVR method solve two ε -SVR problems using the lower bound and upper bound of the interval data. So that we can achieve two hyperplanes. In the proposed method, it does not just use the averaging of these two hyperplanes as a decision-regressor, but introduces different rules for estimating data. The rest of this paper is organized as follows: Section 2 briefly introduces classical ε -SVR. Section 3 presents the proposed interval support vector regression (ISVR) model. Experimental results on several benchmark datasets and real practical Cardiological dataset are given in Section 4. Some conclusions are drawn in Section 5.

2. ε -SVR

Assume that $S=\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_l, y_l)\}$ be a set of l training samples, $\mathbf{x}_i \in \mathbb{R}^m$ is a m -dimensional sample in the input space, $y_i \in \mathbb{R}$, and b is the bias that is a scalar. We are going to seek a linear regression function in the form of $y = \mathbf{w} \cdot \mathbf{x} + b$ where the parameters \mathbf{w} and b are determined as the solution of the following quadratic optimization problem:

$$\begin{aligned} \min_{\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\xi}^*} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ \text{s.t.} \quad & y_i - (\mathbf{w} \cdot \mathbf{x}_i + b) \leq \varepsilon + \xi_i \quad i = 1, 2, \dots, l \\ & \mathbf{w} \cdot \mathbf{x}_i + b - y_i \leq \varepsilon + \xi_i^* \quad i = 1, 2, \dots, l \\ & \xi_i \geq 0, \xi_i^* \geq 0, \quad i = 1, 2, \dots, l. \end{aligned} \tag{2.1}$$

where $\boldsymbol{\xi} = (\xi_1, \dots, \xi_l)^T$ and $\boldsymbol{\xi}^* = (\xi_1^*, \dots, \xi_l^*)^T$ are the vector of slack variables that represent upper and lower constraints on the outputs of the model. $C > 0$ is a regulated positive parameter that controls the trade-off between the error and margin, and ε is input parameters; $\mathbf{y} = (y_1, \dots, y_l)^t$ is a vector of observed values. It is practice that rather than solving the primal problem (2.1), its dual is solved. The Wolfe dual of (2.1) can be written as:

$$\begin{aligned} \min_{\boldsymbol{\alpha}, \boldsymbol{\alpha}^* \in \mathbb{R}^l} \quad & \frac{1}{2} \sum_{i,j=1}^l (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) \mathbf{x}_i^T \mathbf{x}_j + \varepsilon \sum_{i=1}^l (\alpha_i^* + \alpha_i) - \sum_{i=1}^l y_i (\alpha_i^* - \alpha_i) \\ \text{s.t.} \quad & \sum_{i=1}^l (\alpha_i^* - \alpha_i) = 0 \\ & 0 \leq \alpha_i, \alpha_i^* \leq C, \quad i = 1, 2, \dots, l. \end{aligned} \tag{2.2}$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_l)^T$ and $\boldsymbol{\alpha}^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_l^*)^T$ are the vectors of Lagrange multipliers. The main innovation of the proposed method is to model the noisy samples by using interval number in SVR. For this purpose, the data points and penalty coefficient are considered as interval-valued. We present ISVR and reformulate it into an interval quadratic optimization problem. Then, the optimization problem is divided into two quadratic programming problems in a normal way (no interval).

By solving these two quadratic programs, two hyperplanes are obtained which they construct the solution range of the problem. Some classifier rules are given to establish the data points. The experiments show that the proposed method is effective to use for all datasets such as interval-valued, noisy, and normal (no interval).

3. ε -Interval Support vector Regression (ε -ISVR)

In this section, ε -ISVR is introduced. In fact, ε -SVR leads to ε -ISVR when the training data is interval. A notation, “ \sim ”, on a quantity indicates that the quantity has an interval value. Let $\tilde{\mathbf{x}}, \mathbf{x}^L, \mathbf{x}^U \in \mathbb{R}^n$ and $\tilde{A}, A^L, A^U \in \mathbb{R}^{m \times n}$, we say $\mathbf{x}^L \leq \tilde{\mathbf{x}} \leq \mathbf{x}^U$ if $x_i^L \leq \tilde{x}_i \leq x_i^U$, for all $i = 1, 2, \dots, n$, and $A^L \leq \tilde{A} \leq A^U$, if $a_{ij}^L \leq \tilde{a}_{ij} \leq a_{ij}^U$, for all $i = 1, 2, \dots, m$ and $j = 1, 2, \dots, n$, respectively. So we have the following ε -ISVR:

$$\begin{aligned}
 \min_{\mathbf{w}, b, \xi, \xi^*} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\
 \text{s.t.} \quad & y_i - (\mathbf{w} \cdot \tilde{\mathbf{x}}_i + b) \leq \varepsilon + \xi_i \quad i = 1, 2, \dots, l \\
 & \mathbf{w} \cdot \tilde{\mathbf{x}}_i + b - y_i \leq \varepsilon + \xi_i^* \quad i = 1, 2, \dots, l \\
 & \mathbf{x}_i^L \leq \tilde{\mathbf{x}}_i \leq \mathbf{x}_i^U \\
 & \xi_i \geq 0, \xi_i^* \geq 0, \quad i = 1, 2, \dots, l.
 \end{aligned} \tag{3.1}$$

According to the interval value of $\tilde{\mathbf{x}}$, we have an interval quadratic programming problem that the objective function value sets within an interval. Our goal is to find the upper and lower bounds of the range where the target function value is located. Thus, similar to (2.2), we consider the dual of problem (3.1) as follows:

$$\begin{aligned}
 \min_{\boldsymbol{\alpha}, \boldsymbol{\alpha}^* \in \mathbb{R}^l} \quad & \frac{1}{2} \sum_{i,j=1}^l (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) \tilde{\mathbf{x}}_i^T \tilde{\mathbf{x}}_j + \varepsilon \sum_{i=1}^l (\alpha_i^* + \alpha_i) - \sum_{i=1}^l y_i (\alpha_i^* - \alpha_i) \\
 \text{s.t.} \quad & \sum_{i=1}^l (\alpha_i^* - \alpha_i) = 0 \\
 & 0 \leq \alpha_i, \alpha_i^* \leq C, \quad i = 1, 2, \dots, l.
 \end{aligned} \tag{3.2}$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_l)^T$ and $\boldsymbol{\alpha}^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_l^*)^T$ are the vectors of Lagrange multipliers. The matrix form of the QPP (3.2) is follows:

$$\begin{aligned}
 \tilde{Z} = \min_{\boldsymbol{\alpha}, \boldsymbol{\alpha}^* \in \mathbb{R}^l} \quad & \frac{1}{2} \|\tilde{X}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)\|_2^2 + \varepsilon \mathbf{f}^T(\boldsymbol{\alpha} + \boldsymbol{\alpha}^*) - \mathbf{y}^T(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) \\
 \text{s.t.} \quad & \mathbf{f}^T(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) = 0 \\
 & X^L \leq \tilde{X} \leq X^U \\
 & 0 \leq \boldsymbol{\alpha}, \boldsymbol{\alpha}^*,
 \end{aligned} \tag{3.3}$$

where, \tilde{X} is a $m \times l$ matrix with columns $\tilde{\mathbf{x}}_i$, $\mathbf{f} = [1, 1, \dots, 1]^T$ is a $l \times 1$ vector and $\|\tilde{X}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)\|_2^2 = (\tilde{X}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*))^T(\tilde{X}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*))$. In the next section, we introduce the our approach for solving the interval QPP (3.3).

4. Description of the Solving Method

To obtain the interval bounding of objective values, it is sufficient to obtain the lower bound(Z^L) and the upper bound(Z^U) of the objective values of the problem (3.3). The following lemma and theorem help us to determine the values of Z^L and Z^U .

Lemma 4.1. *Let $\tilde{X}, X^L, X^U \in \mathbb{R}^{m \times l}$ such that $X^L \leq \tilde{X} \leq X^U$ and $\alpha \in \mathbb{R}^l$ then*

$$\frac{1}{\sqrt{m}} \|X^L \alpha\|_2 \leq \|\tilde{X} \alpha\|_2 \leq \sqrt{m} \|X^U \alpha\|_2$$

Proof.

$$\begin{aligned} \|\tilde{X} \alpha\|_2 &\leq \|\tilde{X} \alpha\|_1 \\ &\leq \|X^U\|_1 \|\alpha\|_1 \\ &= \max_{\alpha} \|X^U \alpha\|_1 \leq \sqrt{m} \|X^U \alpha\|_2. \\ \|\tilde{X} \alpha\|_2 &\geq \max_{\alpha} \|\tilde{X} \alpha\|_{\infty} \\ &= \|\tilde{X}\|_{\infty} \|\alpha\|_{\infty} \\ &\geq \|X^L\|_{\infty} \|\alpha\|_{\infty} \\ &= \max_{\alpha} \|X^L \alpha\|_{\infty} \\ &\geq \max_{\alpha} \frac{1}{\sqrt{m}} \|X^L \alpha\|_2 \\ &\geq \frac{1}{\sqrt{m}} \|X^L \alpha\|_2. \end{aligned}$$

□

Theorem 4.2. *Let $S = \{\alpha, \alpha^* \in \mathbb{R}^l : f(\alpha - \alpha^*) = 0, \alpha, \alpha^* \geq 0\}$. The optimal value of interval quadratic programming problem (3.3) lies in the interval $[Z^L, Z^U]$ where Z^L and Z^U as follows:*

$$Z^L = \min_{\alpha, \alpha^* \in S} \frac{1}{2} \|X^L(\alpha - \alpha^*)\|_2^2 + \varepsilon f^T(\alpha + \alpha^*) - y^T(\alpha - \alpha^*) \tag{4.1}$$

and

$$Z^U = \min_{\alpha, \alpha^* \in S} \frac{1}{2} \|X^U(\alpha - \alpha^*)\|_2^2 + \varepsilon f^T(\alpha + \alpha^*) - y^T(\alpha - \alpha^*) \tag{4.2}$$

Proof. Based on the Lemma 4.1 and the definition of S , the interval QPP (3.3) takes its lowest value, which has the least objective function and takes its maximum value, which has the largest objective function. Thus, the interval QPP (3.3) is transformed into two QPPs (4.1) and (4.2) to obtain the Z^L and the Z^U . □

Finally, we summarize all the result of this section in the following corollary.

Corollary 4.3. *According to the above theorem, we can transform problems (4.1) and (4.2) into the following standard QPPs:*

$$\begin{aligned}
 Z^L &= \min_{\alpha, \alpha^*} \frac{1}{2} (\alpha - \alpha^*)^T (X^L)^T X^L (\alpha - \alpha^*) + \varepsilon \mathbf{f}^T (\alpha + \alpha^*) - \mathbf{y}^T (\alpha - \alpha^*) \\
 &\text{s.t. } \alpha, \alpha^* \in S
 \end{aligned}
 \tag{4.3}$$

and

$$\begin{aligned}
 Z^U &= \min_{\alpha, \alpha^*} \frac{1}{2} (\alpha - \alpha^*)^T (X^U)^T X^U (\alpha - \alpha^*) + \varepsilon \mathbf{f}^T (\alpha + \alpha^*) - \mathbf{y}^T (\alpha - \alpha^*) \\
 &\text{s.t. } \alpha, \alpha^* \in S
 \end{aligned}
 \tag{4.4}$$

As we have seen, we proved that to solve the regression problem (3.1) with bounded data, we need to solve two problems of regression (4.3) and (4.4) where the problem (4.3) is considered lower bound of data, and the upper bound of data is considered in the problem (4.4). To combine two solutions of the problem (4.3) and (4.4), we conclude that the objective value of the interval quadratic programming (3.1) lies in the range of $[Z^L, Z^U]$. Now, by using the solutions α^L, α^{*L} and α^U, α^{*U} , of the problems (4.3) and (4.4), we can obtain two hyperplanes $h_1(\mathbf{x}) = \mathbf{w}^L \cdot \mathbf{x} + b^L$ and $h_2(\mathbf{x}) = \mathbf{w}^U \cdot \mathbf{x} + b^U$ where

$$\begin{aligned}
 \mathbf{w}^L &= \sum_{i=1}^l (\alpha_i^L - \alpha_i^{*L}) \mathbf{x}_i \\
 b^L &= y_i - \mathbf{w}^L \mathbf{x}_i + \varepsilon, \quad \text{for some } i: \alpha_i^L > 0 \\
 \mathbf{w}^U &= \sum_{i=1}^l (\alpha_i^U - \alpha_i^{*U}) \mathbf{x}_i \\
 b^U &= y_i - \mathbf{w}^U \mathbf{x}_i + \varepsilon, \quad \text{for some } i: \alpha_i^U > 0.
 \end{aligned}$$

Note that, in the case $X^L = X^U = X$, two problems (4.3) and (4.4) are equivalent to the problem (2.1).

5. Decision rules

In this section, some rules for the estimated regressors are given. In fact, we give some rules for specifying the data to appropriate estimated. Once the problems (4.3) and (4.4), are solved, two hyperplanes $h_1(\mathbf{x}) = \mathbf{w}^L \cdot \mathbf{x} + b^L$ and $h_2(\mathbf{x}) = \mathbf{w}^U \cdot \mathbf{x} + b^U$ are obtained. Then, the estimated regressor of the presented method is denoted by $h(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b$. For a better understanding, consider the one-dimensional syntactic interval data with the small ε along with the two estimated regressors $h_1(\mathbf{x})$ and $h_2(\mathbf{x})$ in Figure 2. In the test phase, let an arbitrary sample \mathbf{x}_t is given. This sample is shown in the figure. We denote $\hat{\mathbf{y}}_t$ as the predicted of \mathbf{x}_t and is obtained of the following decision rules:

1. ISVR(Avg.)

$$\hat{\mathbf{y}}_t^{(Avg)} = \frac{1}{2} (h_1(\mathbf{x}_t) + h_2(\mathbf{x}_t)).$$

2. ISVR(Min.)

$$\hat{\mathbf{y}}_t^{(Min)} = h_k(\mathbf{x}_t) \quad \text{where } k = \text{arg min}_{1,2} \left\{ \frac{|h_1(\mathbf{x}_t)|}{\|\mathbf{w}^L\|_2}, \frac{|h_2(\mathbf{x}_t)|}{\|\mathbf{w}^U\|_2} \right\}.$$

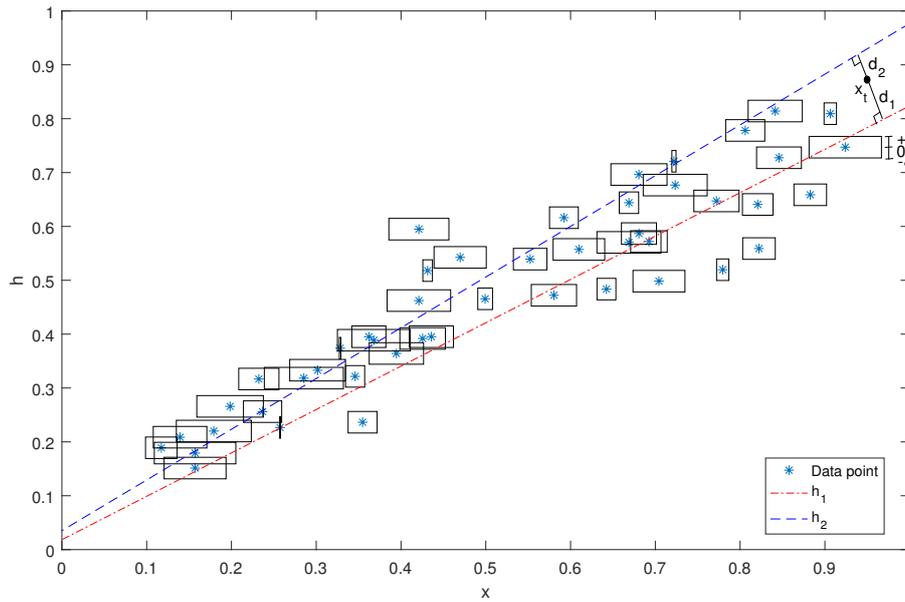


Figure 2: A simple example of ISVR on interval syntactic data

3. ISVR(Max.)

$$\hat{y}_t^{(Max)} = h_k(\mathbf{x}_t) \quad \text{where } k = \arg \max_{1,2} \left\{ \frac{|h_1(\mathbf{x}_t)|}{\|\mathbf{w}^L\|_2}, \frac{|h_2(\mathbf{x}_t)|}{\|\mathbf{w}^U\|_2} \right\}.$$

4. ISVR(Weighted1)

$$\hat{y}_t^{(W1)} = \frac{d_1 h_1(\mathbf{x}_t) + d_2 h_2(\mathbf{x}_t)}{d_1 + d_2} \quad \text{where } d_1 = \frac{|h_1(\mathbf{x}_t)|}{\|\mathbf{w}^L\|_2} \text{ and } d_2 = \frac{|h_2(\mathbf{x}_t)|}{\|\mathbf{w}^U\|_2}.$$

5. ISVR(Weighted2)

$$\hat{y}_t^{(W2)} = \frac{d_2 h_1(\mathbf{x}_t) + d_1 h_2(\mathbf{x}_t)}{d_1 + d_2} \quad \text{where } d_1 = \frac{|h_1(\mathbf{x}_t)|}{\|\mathbf{w}^L\|_2} \text{ and } d_2 = \frac{|h_2(\mathbf{x}_t)|}{\|\mathbf{w}^U\|_2}.$$

For sample \mathbf{x}_t , the distance d_1 and d_2 are represented in figure 2.

6. Numerical experiments

In this section, some experiments are constructed to illustrate the performance of ISVR compare with the mentioned methods on benchmark and real practical interval datasets. All computations have been performed with symbolic computation software MATLAB, and the calculations are implemented on a machine with Intel Core 5 Duo processor 2 GHz and 4 GB RAM. Note that, all datasets are normalized in the range $[0, \infty)$.

Evaluation criteria

In order to measure the efficiency of the proposed algorithm, some evaluation criteria customarily used should be stated firstly. we assume l be the number of testing samples. \hat{y}_i as the predict value

of y_i and $\bar{y}_i = \frac{1}{l} \sum_{i=1}^l y_i$. Then the definitions of metrics and their calculations are as follow:

$$\begin{aligned}
 SSE &= \sum_{i=1}^l (y_i - \hat{y}_i)^2 & SST &= \sum_{i=1}^l (y_i - \bar{y}_i)^2 & SSR &= \sum_{i=1}^l (\bar{y}_i - \hat{y}_i)^2 \\
 MAPE &= \frac{1}{l} \sum_{i=1}^l \left| \frac{y_i - \hat{y}_i}{y_i} \right| & NMSE &= \frac{SSE}{SST} & R^2 &= \frac{SSR}{SST}
 \end{aligned}$$

Table 1: Properties of UCI benchmark datasets for Example 1.

Dataset	No. samples	Attribute
Energy efficiency	768	8
ISE	536	8
conc com str	1030	8
Bodyfat	252	14
Yacht_H	308	6
eumite2001	367	16
AM	392	7
Housing	506	13
triazines	186	60
Servo	167	4
Cpusmall	8,192	12
Cadata	20,640	8
Abalone	4,177	8

Example 1: Standard Datasets

We use standard UCI datasets [4] that the properties of these datasets are summarized in Table 1. Also, we describe the attribute information for some datasets as follows:

Istanbul Stock Exchange(ISE): This dataset includes returns of Istanbul Stock Exchange with seven other international index; SP, DAX, FTSE, NIKKEI, BOVESPA, MSCE_EU, MSCILEM from Jun 5, 2009 to Feb 22, 2011.

Auto MPG(AM): The data be about fuel consumption according to cylinders, displacement, horsepower, weight, acceleration, model year and origin of the automobile.

Energy efficiency: This data regraded into assigning the heating load and cooling load requirements of buildings (that is, energy efficiency) as a function of Relative Compactness, Surface Area, Wall Area, Roof Area, Overall Height, Orientation, Glazing Area and Glazing Area Distribution, as building parameters.

Servo: It is about predicting for the rise time of a servomechanism subject to gain settings and choices of mechanical connections.

Yacht Hydrodynamics(Yacht_H): This data set used to predict the hydrodynamic performance of sailing yachts from dimensions and velocity.

Abalone: Predicting the age of abalone from the measurement of physical parameters such as Shell weight, Sex, Length, Diameter, Height, Whole weight, Shucked weight, and Viscera weight.

Concrete Compressive Strength(CCS): The evaluation of concrete compressive strength by the measure of Cement, Blast Furnace Slag, Fly Ash, Water, Superplasticizer, Coarse Aggregate, Fine Aggregate, Age. In order to evaluate the results of ISVR, we compared it with some variants of LibLinear-SVR (LibSVR), L2LibSVR, L1LibSVR, L2PrimalLibSVR [7], PINSVR[16] and

Table 2: Comparative results for Example 2

Dataset	Regressor	NMSE	R^2	MAPE	CPU Sec.
Bodyfat	L2LibSVR	0.09930 ± 0.05672	0.92966 ± 0.14590	7.99165 ± 1.70334	0.0135 ± 0.00110
	L1LibSVR	0.09930 ± 0.05672	0.92966 ± 0.14590	7.99165 ± 1.70334	0.0118 ± 0.00147
	L2PrimalLibSVR	0.09930 ± 0.05672	0.92966 ± 0.14590	7.99165 ± 1.70334	0.0116 ± 0.00174
	ISVR(max)	0.11491 ± 0.06700	0.93157 ± 0.20187	8.35196 ± 2.52253	0.0072 ± 0.00312
	PINSVR-Linear	0.46586 ± 0.12625	0.30195 ± 0.13743	18.23003 ± 2.50091	3.8176 ± 0.30513
	ETSVR-Linear	0.09765 ± 0.05700	0.95215 ± 0.14344	7.86052 ± 1.68094	0.0160 ± 0.00383
Yacht_H	L2LibSVR	0.04717 ± 0.01749	0.93429 ± 0.07644	19.39353 ± 6.89909	0.0125 ± 0.00089
	L1LibSVR	0.04717 ± 0.01749	0.93429 ± 0.07644	19.39353 ± 6.89909	0.0095 ± 0.00050
	L2PrimalLibSVR	0.04717 ± 0.01749	0.93429 ± 0.07644	19.39353 ± 6.89909	0.0095 ± 0.00061
	ISVR(max)	0.04612 ± 0.01758	0.98215 ± 0.06637	21.57699 ± 7.86366	0.0075 ± 0.00220
	PINSVR-Linear	0.11289 ± 0.02417	0.53704 ± 0.03947	64.70078 ± 47.87316	6.4480 ± 0.78911
	ETSVR-Linear	0.05478 ± 0.02454	0.94569 ± 0.07541	23.87207 ± 16.17364	0.0262 ± 0.01283
Servo	L2LibSVR	0.43676 ± 0.16746	0.64465 ± 0.20863	99.74835 ± 43.92444	0.0265 ± 0.06859
	L1LibSVR	0.43676 ± 0.16746	0.64465 ± 0.20863	99.74835 ± 43.92444	0.0041 ± 0.00077
	L2PrimalLibSVR	0.43676 ± 0.16746	0.64465 ± 0.20863	99.74835 ± 43.92444	0.0042 ± 0.00069
	ISVR(max)	0.45329 ± 0.22762	0.75423 ± 0.25225	94.81609 ± 37.57911	0.0039 ± 0.00113
	PINSVR-Linear	0.64992 ± 0.07863	0.18087 ± 0.09254	116.33048 ± 60.74280	0.8807 ± 0.51268
	ETSVR-Linear	0.43823 ± 0.17011	0.65816 ± 0.20948	99.76264 ± 43.86762	0.0185 ± 0.02386
Cadata	L2LibSVR	0.42264 ± 0.01946	0.57654 ± 0.02560	39.73711 ± 1.92486	1.5469 ± 0.38857
	L1LibSVR	0.42264 ± 0.01946	0.57654 ± 0.02560	39.73711 ± 1.92486	1.5503 ± 0.23494
	L2PrimalLibSVR	0.42264 ± 0.01946	0.57654 ± 0.02560	39.73711 ± 1.92486	1.9064 ± 0.57790
	ISVR(max)	0.44690 ± 0.02216	0.73164 ± 0.03135	40.80747 ± 1.33781	1.6657 ± 0.69536
	PINSVR-Linear	-	-	-	-
	ETSVR-Linear	-	-	-	-
triazines	L2LibSVR	1.33812 ± 0.48467	0.91103 ± 0.51637	109.86739 ± 251.92113	0.0194 ± 0.00192
	L1LibSVR	1.33812 ± 0.48467	0.91103 ± 0.51637	109.86739 ± 251.92113	0.0181 ± 0.00200
	L2PrimalLibSVR	1.33812 ± 0.48467	0.91103 ± 0.51637	109.86739 ± 251.92113	0.0190 ± 0.00310
	ISVR(max)	1.14714 ± 0.42146	0.75809 ± 0.32253	118.02831 ± 285.11074	0.0186 ± 0.00516
	PINSVR-Linear	1.19906 ± 0.70085	0.62107 ± 0.56471	83.60975 ± 167.11965	1.2541 ± 0.39534
	ETSVR-Linear	2.03879 ± 1.55145	0.58868 ± 1.30276	113.05064 ± 253.08299	0.0117 ± 0.00192
CCS	L2LibSVR	0.45114 ± 0.05373	0.55935 ± 0.05960	130.07296 ± 72.28089	0.0360 ± 0.00605
	L1LibSVR	0.45114 ± 0.05373	0.55935 ± 0.05960	130.07296 ± 72.28089	0.0334 ± 0.01084
	L2PrimalLibSVR	0.45114 ± 0.05373	0.55935 ± 0.05960	130.07296 ± 72.28089	0.0306 ± 0.00390
	ISVR(min)	0.45641 ± 0.06492	0.62638 ± 0.06130	125.36786 ± 70.12245	0.0302 ± 0.00532
	PINSVR-Linear	0.47915 ± 0.04889	0.34947 ± 0.03278	157.98167 ± 98.90301	550.6574 ± 676.76861
	ETSVR-Linear	0.45209 ± 0.05309	0.56710 ± 0.05935	129.06776 ± 71.46800	0.4143 ± 0.24584
eunite2001	L2LibSVR	0.20663 ± 0.04264	0.84448 ± 0.17765	16.52018 ± 6.77425	0.0192 ± 0.00181
	L1LibSVR	0.20663 ± 0.04264	0.84448 ± 0.17765	16.52018 ± 6.77425	0.0187 ± 0.00388
	L2PrimalLibSVR	0.20663 ± 0.04264	0.84448 ± 0.17765	16.52018 ± 6.77425	0.0184 ± 0.00257
	ISVR(max)	0.21403 ± 0.04592	0.90000 ± 0.17796	15.82318 ± 5.92828	0.0175 ± 0.00485
	PINSVR-Linear	0.29799 ± 0.05995	0.46388 ± 0.09109	25.10015 ± 13.06550	12.3530 ± 1.31250
	ETSVR-Linear	0.20613 ± 0.04389	0.84835 ± 0.17744	16.42712 ± 6.68658	0.0253 ± 0.00573
Housing	L2LibSVR	0.32431 ± 0.07195	0.75273 ± 0.18406	32.53561 ± 14.02016	0.0387 ± 0.00652
	L1LibSVR	0.32431 ± 0.07195	0.75273 ± 0.18406	32.53561 ± 14.02016	0.0347 ± 0.00415
	L2PrimalLibSVR	0.32431 ± 0.07195	0.75273 ± 0.18406	32.53561 ± 14.02016	0.0370 ± 0.00335
	ISVR(max)	0.32821 ± 0.07747	0.82283 ± 0.18607	32.55391 ± 13.08202	0.0380 ± 0.00868
	ETWSVR	0.87079 ± 0.22073	0.44428 ± 0.11143	34.02766 ± 13.41165	0.2378 ± 0.03842
	PINSVR-Linear	0.38714 ± 0.06232	0.45320 ± 0.10970	32.37592 ± 12.05503	40.8221 ± 3.84491
Abalone	L2LibSVR	0.54117 ± 0.04314	0.47407 ± 0.06956	19.79701 ± 0.64086	0.1861 ± 0.02603
	L1LibSVR	0.54117 ± 0.04314	0.47407 ± 0.06956	19.79701 ± 0.64086	0.1734 ± 0.01463
	L2PrimalLibSVR	0.54117 ± 0.04314	0.47407 ± 0.06956	19.79701 ± 0.64086	0.1679 ± 0.02340
	ISVR(max)	0.64559 ± 0.07809	0.72338 ± 0.12473	21.64475 ± 0.61536	0.1717 ± 0.01786
	ETSVR-Linear	0.54064 ± 0.04388	0.47647 ± 0.06828	19.44753 ± 0.58144	4.3512 ± 0.16970
	PINSVR-Linear	-	-	-	-
ISE	L2LibSVR	0.32490 ± 0.06647	0.70168 ± 0.18690	106.26871 ± 111.04793	0.0466 ± 0.07059
	L1LibSVR	0.32490 ± 0.06647	0.70168 ± 0.18690	106.26871 ± 111.04793	0.0202 ± 0.00322
	L2PrimalLibSVR	0.32490 ± 0.06647	0.70168 ± 0.18690	106.26871 ± 111.04793	0.0209 ± 0.00243
	ISVR(min)	0.38689 ± 0.10663	0.99824 ± 0.29673	129.35729 ± 149.74788	0.0220 ± 0.00727
	PINSVR-Linear	0.47654 ± 0.05557	0.26617 ± 0.05977	112.40933 ± 46.38591	48.7498 ± 5.32858
	ETSVR-Linear	0.32719 ± 0.06551	0.71475 ± 0.18752	106.64141 ± 113.34048	0.1382 ± 0.07375

Table 3: Comparison of L2LibSVR, L1LibSVR and L2PrimalLibSVR methods to ISVR according to different criteria on some datasets.

Dataset	Regressor	NMSE	R^2
AM	L2LibSVR	0.00833 ± 0.00298	0.82422 ± 0.05669
	L1LibSVR	0.00934 ± 0.00432	0.82041 ± 0.05758
	L2PrimalLibSVR	0.01112 ± 0.00389	0.76601 ± 0.07834
	ISVR(avg)	0.01281 ± 0.00406	0.72944 ± 0.07816
	ISVR(min)	0.01372 ± 0.00394	0.71108 ± 0.08272
	ISVR(max)	0.01244 ± 0.00415	0.74141 ± 0.07504
	ISVR(weighted1)	0.01286 ± 0.00407	0.72699 ± 0.07849
	ISVR(weighted2)	0.01281 ± 0.00402	0.73131 ± 0.07772
Energy efficiency	L2LibSVR	0.14896 ± 0.03216	0.88900 ± 0.08438
	L1LibSVR	0.14896 ± 0.03216	0.88900 ± 0.08438
	L2PrimalLibSVR	0.14896 ± 0.03216	0.88900 ± 0.08438
	ISVR(avg)	0.15493 ± 0.03211	0.92306 ± 0.07988
	ISVR(min)	0.15426 ± 0.03208	0.89784 ± 0.07747
	ISVR(max)	0.15600 ± 0.03218	0.94868 ± 0.08235
	ISVR(weighted1)	0.15493 ± 0.03211	0.92235 ± 0.07984
	ISVR(weighted2)	0.15492 ± 0.03211	0.92377 ± 0.07991
ISE	L2LibSVR	0.32490 ± 0.06647	0.70168 ± 0.18690
	L1LibSVR	0.32490 ± 0.06647	0.70168 ± 0.18690
	L2PrimalLibSVR	0.32490 ± 0.06647	0.70168 ± 0.18690
	ISVR(avg)	0.38064 ± 0.09468	0.96615 ± 0.28407
	ISVR(min)	0.38689 ± 0.10663	0.99824 ± 0.29673
	ISVR(max)	0.38076 ± 0.08384	0.93692 ± 0.27190
	ISVR(weighted1)	0.38060 ± 0.09471	0.96586 ± 0.28385
	ISVR(weighted2)	0.38069 ± 0.09464	0.96646 ± 0.28429
CPUsmall	L2LibSVR	0.52156 ± 0.07547	0.70571 ± 0.17408
	L1LibSVR	0.52156 ± 0.07547	0.70571 ± 0.17408
	L2PrimalLibSVR	0.52156 ± 0.07547	0.70571 ± 0.17408
	ISVR(avg)	0.66254 ± 0.08943	0.55077 ± 0.16938
	ISVR(min)	0.66359 ± 0.08828	0.57282 ± 0.16985
	ISVR(max)	0.66604 ± 0.09100	0.52416 ± 0.16931
	ISVR(weighted1)	0.66258 ± 0.08942	0.55089 ± 0.16942
	ISVR(weighted2)	0.66249 ± 0.08943	0.55065 ± 0.16935

Table 4: Cardiological interval data set of example 2

n	Pulse rate	Systolic blood pressure	Diastolic
1	[44-68]	[90-100]	[50-70]
2	[60-72]	[90-130]	[70-90]
3	[56-90]	[140-180]	[90-100]
4	[70-112]	[110-142]	[80-108]
5	[54-72]	[90-100]	[50-70]
6	[70-100]	[130-160]	[80-110]
7	[63-75]	[60-100]	[140-150]
8	[72-100]	[130-160]	[76-90]
9	[76-98]	[110-190]	[70-110]
10	[86-96]	[138-180]	[90-110]
11	[86-100]	[110-150]	[78-100]

ETSVR[15]. Among the these algorithms, PINSVR and ETSVR are determined a pair of nonparallel proximal functions. Also, we apply the decision rules in Section 5, and in term of using decision rules, the best result is reported in Table 2. In addition, the interval-valued data are constructed as follows. The lower case is obtained as reducing $\alpha_1 = 0.02\%$ and the upper case is obtained as increasing $\alpha_2 = 0.03\%$ from data (i.e., $X^L = X - \alpha_1 X$ and $X^U = X + \alpha_2 X$). The Results of 10-fold cross-validation includes mean, standard deviation (mean \pm std) and the average of runtime (sec) are reported in Table 2. PINSVR and ETSVR algorithms are not evaluated in some datasets with sample size more than 1,500, since these algorithms is not scalable. The comparison results show that ISVR achieves the best or closest to the best result for all of the datasets. To evaluate the decision rules, for some dataset, the results is reported for all of the decision rules in Table 3. The results in Tables 2 and 3 show that there is no significant difference between the use of these rules.

Example 2: Interval Data Points

For further evaluation, a real practical Cardiological dataset [9] is test in this example. The Cardiological dataset concerns the record of the pulse rate, the systolic blood pressure and diastolic blood pressure. In Table 4, information of the eleven patients is reported. The aim is to obtain a linear regression model such as $y = a_0 + a_1x_1 + a_2x_2$ to predict the interval values \hat{y} of y (The pulse rate) where x_1 (Systolic blood pressure) and x_2 (Diastolic blood pressure) are interval values.

The CM([3]), CRM([2]), CCRM([11]), and ISVR methods were applied on the Cardiological interval data set and the following linear regression equations were obtained as:

$$\begin{aligned}
 \text{CM :} & & y_L &= 21.17 + 0.33x_1 + 0.17x_2 & \text{and} & & y_U &= 21.17 + 0.33x_1 + 0.17x_2, \\
 \text{CRM :} & & y_L &= y^c - y^r & \text{and} & & y_U &= y^c + y^r, \\
 & \text{where} & y^c &= 21.17 + 0.33x_1^c + 0.17x_2^c & \text{and} & & y^r &= 20.13 - 0.15x_1^r + 0.35x_2^r, \\
 \text{CCRM :} & & y_L &= y^c - y^r & \text{and} & & y_U &= y^c + y^r, \\
 & \text{where} & y^c &= 21.17 + 0.33x_1^c + 0.17x_2^c & \text{and} & & y^r &= 17.96 + 0.20x_2^r, \\
 \text{ISVR :} & & y_L &= 0.12 + 0.21x_1 + 0.60x_2 & \text{and} & & y_U &= 0.002 + 0.59x_1 + 0.06x_2.
 \end{aligned}$$

where y^c and y^r are obtained as the prediction of y such that the center and the range of interval value data are considered for the regression linear, respectively. We compare the proposed method with CM, CRM and CCRM methods on this dataset. Table 5 show the prediction of the pulse rate for the eleven patients. Also, the comparison of the mentioned methods using the following errors is reported in Table 6.

Table 5: The evaluation of ISVR compared to CM, CRM and CCRM through the calculation the interval values of Pulse variable.

Pulse	ISVR	CM	CRM	CCRM
[46-68]	[49 -63]	[59 -66]	[37- 88]	[41-85]
[60 - 72]	[61- 82]	[63- 79]	[50- 92]	[49-93]
[56-90]	[84-112]	[83-98]	[72-108]	[70-110]
[70-112]	[71-90]	[71-86]	[54-104]	[55-102]
[54-72]	[49-63]	[59-66]	[37-88]	[41- 85]
[70 -100]	[75-101]	[78-93]	[59-111]	[61-109]
[63-75]	[66-95]	[78-88]	[50-115]	[57-109]
[72-100]	[73-100]	[77-89]	[63-104]	[62-104]
[76-98]	[65-119]	[69-103]	[64-108]	[60-112]
[86-96]	[83-113]	[82-99]	[70-111]	[69-113]
[86-100]	[70-95]	[71-88]	[57-101]	[57-102]

Table 6: The evaluation of ISVR compared to CM, CRM and CCRM through the calculation of $RMSE_L, RMSE_U, R_L^2$ and R_U^2

Method	$RMSE_L$	$RMSE_U$	R_L^2	R_U^2
CM	11.09	10.41	0.5774	0.6932
CRM	9.81	8.94	12.8457	0.9329
CCRM	9.73	9.17	13.2450	0.9329
ISVR	9.63	7.29	0.8229	1.6326

$$\begin{aligned}
 RMSE_L &= \sqrt{\frac{\sum_{i=1}^n (y_L^{(i)} - \hat{y}_L^{(i)})^2}{n}}, & R_L^2 &= \frac{\sum_{i=1}^n (\bar{y}_L^{(i)} - \hat{y}_L^{(i)})^2}{\sum_{i=1}^n (y_L^{(i)} - \bar{y}_L^{(i)})^2} \\
 RMSE_U &= \sqrt{\frac{\sum_{i=1}^n (y_U^{(i)} - \hat{y}_U^{(i)})^2}{n}}, & R_U^2 &= \frac{\sum_{i=1}^n (\bar{y}_U^{(i)} - \hat{y}_U^{(i)})^2}{\sum_{i=1}^n (y_U^{(i)} - \bar{y}_U^{(i)})^2}
 \end{aligned}$$

From Table 5 and 6, we deduce that the ISVR method performs better than the CM, CRM and CCRM methods in this data set.

7. Conclusion

In this paper, we investigated SVR problem for interval dataset. The proposed method named Interval Support Vector Regression (ISVR) and, ISVR reformulated the problem into two SVR problems according to the lower and upper bound of data. We obtained two hyperplanes, each of which was based on the lower and upper of data, respectively. To predict interval data, ISVR. To predict the test data, several decision rules were introduced. In contrast to the existing methods, ISVR did not just use the averaging of these two hyperplanes as a decision-regressor but it introduced different decision rules for estimating data. The experimental results showed that the proposed approach is effective on interval data and those achieve a better performance on real-world datasets than the existing methods.

References

- [1] J. Bi and K. P. Bennett, *A geometric approach to support vector regression*, Neurocomp. 55 (1-2) (2003) 79–108.
- [2] L. Billard, *Dependencies and Variation Components of Symbolic Interval-Valued Data*, Selected Contributions in Data Analysis and Classification, Springer, (2007) 3–12.
- [3] L. Billard and E. Diday, *Regression Analysis for Interval-Valued Data*, Data Analysis, Classification, and Related Methods, Springer, (2000) 369–374.
- [4] C. Blake, *Uci repository of machine learning databases*, <http://www.ics.uci.edu/~mlearn/MLRepository.html>. 1998.
- [5] C. J. Burges, *A tutorial on support vector machines for pattern recognition*, Data Min. Knowl- edge Disc. 2(2) (1998) 121–167.
- [6] Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine learning 20 (3), 273–297.
- [7] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang and C.-J. Lin, *Liblinear: A library for large linear classification*, J. Mach. Lear. Res. 9 (2008) 1871–1874.
- [8] E. d. A. L. Neto, G. M. Cordeiro and F. d. A. de Carvalho, *Bivariate symbolic regression models for interval-valued variables*, J. Stat. Comput. Simul. 81(11) (2011) 1727–1744.
- [9] E. d. A. L. Neto, F. A. de Carvalho and C. P. Tenorio, *Univariate and multivariate linear regression methods to predict interval-valued features*, Aust. Joint Conf. Artif. Intel., Springer, (2004) 526–537.
- [10] E. d. A. L. Neto and F. A. de Carvalho, *Centre and range method for fitting a linear regression model to symbolic interval data*, Comput. Stat. Data Anal. 52(3) (2008) 1500–1515.
- [11] E. d. A. L. Neto and F. A. de Carvalho, *Constrained linear regression models for symbolic interval valued variables*, Comput. Stat. Data Anal. 54 (2) (2010) 333–347.
- [12] X. Peng, *Efficient twin parametric insensitive support vector regression model*, Neurocomp. 79 (2012) 26–38.
- [13] X. Peng and Y. Wang, *The robust and efficient adaptive normal direction support vector regression*, Expert Syst. Appl. 38(4) (2011) 2998–3008.
- [14] Y.-H. Shao, C.-H. Zhang, Z.-M. Yang, L. Jing and N.-Y. Deng, *An ε -twin support vector machine for regression*, Neural Comput. Appl. 23(1) (2013) 175–185.
- [15] Shao, Y.-H., Zhang, C.-H., Yang, Z.-M., Jing, L., Deng, N.-Y., 2013b. An ε -twin support vector machine for regression. Neural Computing and Applications 23 (1), 175–185.
- [16] Z.-M. Yang, X.-Y. Hua, Y.-H. Shao and Y.-F. Ye, *A novel parametric-insensitive nonparallel support vector machine for regression*, Neurocomp. 171 (2016) 649–663.