Interval Support Vector Machine in Regression Analysis

Ameneh Arjmandzadeh[1], Sohrab Effati[2], Mohammad Zamirian[3*]
[1] Department of Mathematics, Islamic Azad University of Bojnourd Branch, Bojnourd, Iran
a.arjmandzadeh@gmail.com
[2] Department of Mathematics, Ferdowsi University of Mashhad, Mashhad, Iran
effati911@yahoo.com
[3] Department of Mathematics, Islamic Azad University of Bojnourd Branch, Bojnourd, Iran
zamirianm@yahoo.com

Abstract

Support vector machines (SVMs) have been widely applied in regression analysis. In this paper, the application of SVM in regression for interval samples is proposed. The standard support vector regression (SVR), is a quadratic optimization problem that is formulated according to the form of training samples and optimal hyperplane is obtained. In real world, the parameters are seldom known and usually are estimated. In this paper we propose an interval support vector regression (ISVR) problem which the training samples are interval values. Using duality theorem and applying variable transformation theorem the problem is solved and two hyperplanes correspond to the upper bound and the lower bound of solution set is obtained. Efficiency of our approach is confirmed by a numerical example.

Keywords: Support vector machine, Regression analysis, Interval quadratic optimization problem.

1. Introduction

Recently, a novel machine learning technique, is called SVM, drawn much attention in the fields of pattern classification and regression forecasting. SVM was first introduced by Vapnik in 1995 [2]. SVM is a kind of classifier's studing method on statistic study theory.

---

[*]Corresponding author: Mohammad Zamirian (zamirianm@yahoo.com)

This algorithm derives from linear classifier, and can solve the problem of two kind classifier, later this algorithm applies in non-linear fields, that is to say, we can find the optimal hyperplane (large margin) to classify the samples set. It is an approximate implementation to the structure risk minimization method [9]. SVM has been very successful in pattern recognition and function estimation problems. Linear regression models are widely used today in business, administration, economics, engineering, as well as in many other traditionally no quantitative fields such as social, health, and biological sciences.

For the last few years, there has been an increasing interest in fuzzy systems which incorporate tools well-known from the statistical learning theory. Fuzzy clustering with a weighted (or fuzzy) ε-insensitive loss function is introduced in [3, 4, 5]. The above method leads to improved robustness of outliers with respect to the traditional fuzzy clustering methods. The support vector fuzzy regression machines have been introduced in [8]. The support vector interval regression network has been established in [6]. A differentiable approximation of the misclassification rate with the use of the empirical risk minimization principle to improve learning of a neuro-fuzzy classifier is proposed in [7].

In all of references, the training samples are proposed as crisp points. But, in fact, these samples are undetermined. Then we suppose the samples as interval values which resulting in an interval quadratic optimization problem. A numerical solution to interval quadratic programming was described by Liu and Wang (2007) [1]. In the proposed method by Liu and Wang since the parameters are interval-valued, the objective value is achieved interval-valued by solving a pair of two-level mathematical programs. In this paper we solve the ISVR problem by solving a quadratic program problem.

2. The problem

The SVM can be successfully applied to regression as follows:

Given training data $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n) \in \mathfrak{R}^m \times \mathfrak{R}$ where $x \in \mathfrak{R}^m$ is a data point in input space and $y_i \in \mathfrak{R}$ is the corresponding target of the model output. The goal is to find an approximation function $f(x)$, which has at most ε deviation from the actual target y_i for all the training data. In the ε-SV regression [Vapnik, 1995], any deviation that is less than ε is allowed. Consider a linear function, $f(x) = w^T x + b$ where $w \in \mathfrak{R}^m$ and $b \in \mathfrak{R}$ are unknown. The objective of SVR is to minimize the Euclidean norm of w as long as deviation from the actual target y_i is less than ε. This convex optimization problem is shown as follows:

$$Min \quad \frac{1}{2}\|w\|^2$$

$$s.t \quad y_i - w^T x_i - b \leq \varepsilon + \xi_i \quad i = 1,...,n \tag{1}$$

$$w^T x_i + b - y_i \leq \varepsilon + \xi_i' \quad i = 1,...,n$$

The implicit assumption in (1) is that the problem is feasible as the function $f(x)$ that approximates all input data, $\{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$ whit ε deviation exists. To construct a support vector machine for approximating a target, d, the following loss function is used:

$$L_\varepsilon(d, y) = \begin{cases} |d - y| - \varepsilon & \text{for } |d - y| \geq \varepsilon \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Where ε is a predetermined parameter with positive value.

The loss function in (2) is called the ε-insensitive loss function. This loss function implies that this function has a value only when the deviation of output, y, from the desired target, d, is greater than the deviation parameter, ε. It is graphically presented as shown in figure 1.

Figure 1. The ε-insensitive loss function
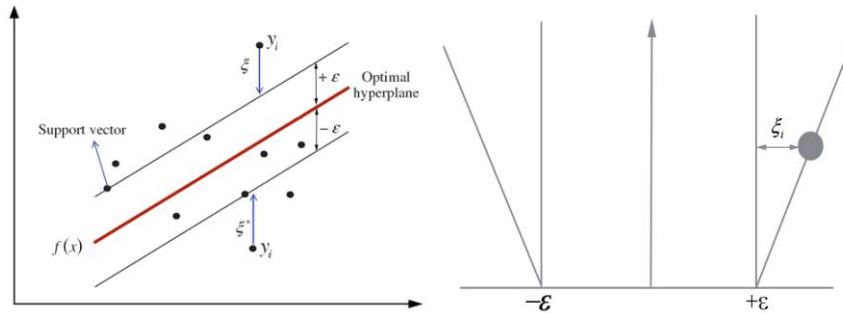
To deal whit the feasible constraints of the optimization problem (1), slack variables, ξ_i, ξ_i', are introduced to form the following problem:

$$Min \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i')$$

$$s.t \quad y_i - w^T x_i - b \leq \varepsilon + \xi_i \qquad i = 1,...,n \tag{3}$$

$$w^T x_i + b - y_i \leq \varepsilon + \xi_i' \qquad i = 1,...,n$$

$$\xi_i, \xi_i' \geq 0 \qquad i = 1,...,n$$

Where C is a given large real number as the coefficient of penalty.

In this paper, we suppose that (x_i, y_i), $i = 1,...,n$, are interval samples named $(\tilde{x}_i, \tilde{y}_i) = ([x_i^l, x_i^u], [y_i^l, y_i^u])$, $i = 1,...,n$. Now the goal is to find the approximation function $f(\tilde{x}) = \tilde{w}^T x + b$ which has at most ε deviation from the actual target y_i for all the training data. Then the problem (3) is converted to the following problem:

$$Min \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i')$$

$$s.t \quad \tilde{y}_i - w^T \tilde{x}_i - b \leq \varepsilon + \xi_i \qquad i = 1,...,n$$

$$w^T \tilde{x}_i + b - \tilde{y}_i \leq \varepsilon + \xi_i' \qquad i = 1,...,n \tag{4}$$

$$\xi_i, \xi_i' \geq 0 \qquad i = 1,...,n$$

$$x_i^l \le \tilde{x}_i \le x_i^u \qquad\qquad i = 1,...,n$$

$$y_i^l \le \tilde{y}_i \le y_i^u \qquad\qquad i = 1,...,n$$

Where $x_i^l = (x_{1i}^l,..., x_{mi}^l)$, $x_i^u = (x_{1i}^u,..., x_{mi}^u)$, y_i^l and y_i^u are belong to $\Re$, also $\tilde{x}_i = [x_i^l, x_i^u] = ([x_{1i}^l, x_{1i}^u],...[x_{mi}^l, x_{mi}^u])$ and define product operation between two one-dimensional intervals $\tilde{x} = [x^l, x^u]$ and $\tilde{y} = [y^l, y^u]$ such as: $\tilde{x}\tilde{y} = [(xy)^l, (xy)^u]$ where $(xy)^l$ and $(xy)^u$ are achieved by solving the following two optimization problems:

$$(xy)^l = \min \quad \tilde{x}\tilde{y} \qquad\qquad\qquad (xy)^u = \max \quad \tilde{x}\tilde{y}$$

$$s.t. \quad x^l \le \tilde{x} \le x^u \qquad \text{and} \qquad s.t. \quad x^l \le \tilde{x} \le x^u$$

$$y^l \le \tilde{y} \le y^u \qquad\qquad\qquad\qquad y^l \le \tilde{y} \le y^u$$

And we define $\tilde{x} + \tilde{y} = [x^l + y^l, x^u + y^u]$.

The Lagrange function is as follows:

$$L_P = (w,b,\xi_i,\xi_i',\alpha_i,\alpha_i',\beta_i,\beta_i',\gamma_i,\gamma_i',\lambda_i,\lambda_i') = \frac{1}{2}w^T w + C\sum_{i=1}^{n}(\xi_i + \xi_i') - \sum_{i=1}^{n}(\beta_i\xi_i + \beta_i'\xi_i')$$

$$-\sum_{i=1}^{n}\alpha_i[w^T\tilde{x}_i + b - \tilde{y}_i + \varepsilon + \xi_i] - \sum_{i=1}^{n}\alpha_i'[\tilde{y}_i - w^T\tilde{x}_i - b + \varepsilon + \xi_i']$$

$$-\sum_{i=1}^{n}\gamma_i(x_i^u - \tilde{x}_i) - \sum_{i=1}^{n}\gamma_i'(\tilde{x}_i - x_i^l) - \sum_{i=1}^{n}\lambda_i(y_i^u - \tilde{y}_i) - \sum_{i=1}^{n}\lambda_i'(\tilde{y}_i - y_i^l)$$

Thus; a necessary and sufficient condition for minimum the lagrangian function over the variables w, b, is that the gradient vanishes, that is:

$$\begin{cases} \dfrac{\partial L_P}{\partial w} = 0 \\[2mm] \dfrac{\partial L_P}{\partial b} = 0 \\[2mm] \dfrac{\partial L_P}{\partial \xi_i} = 0 \\[2mm] \dfrac{\partial L_P}{\partial \xi_i'} = 0 \end{cases} \quad \text{or} \quad \begin{cases} w = \displaystyle\sum_{i=1}^{n}(\alpha_i - \alpha_i')\tilde{x}_i \\[2mm] \displaystyle\sum_{i=1}^{n}(\alpha_i - \alpha_i') = 0 \\[2mm] \displaystyle\sum_{i=1}^{n}C - \alpha_i = 0 \\[2mm] \displaystyle\sum_{i=1}^{n}C - \alpha_i' = 0 \end{cases} \qquad (5)$$

Therefore; under the above optimality conditions, maximize L_P over $\alpha \ge 0$ is equivalent to solve the following problem:

$$\max \quad F(\alpha, \alpha') = \sum_{i=1}^{n} \tilde{y}_i(\alpha_i - \alpha_i')$$

$$-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i - \alpha_i')(\alpha_j - \alpha_j')\tilde{x}_i^T\tilde{x}_j - \varepsilon\sum_{i=1}^{n}(\alpha_i + \alpha_i')$$

$$s.t. \quad \sum_{i=1}^{n}(\alpha_i - \alpha_i') = 0 \qquad\qquad (6)$$

$$0 \leq \alpha_i \leq C \qquad i = 1, \ldots, n$$

$$0 \leq \alpha_i^{'} \leq C \quad i = 1, \ldots, n$$

$$x_i^l \leq \tilde{x}_i \leq x_i^u \quad i = 1, \ldots, n$$

$$y_i^l \leq y_i \leq y_i^u \quad i = 1, \ldots, n$$

The above problem is a quadratic program with interval coefficients $(\tilde{x}_i^T \tilde{x}_j)$, so we replace the upper and lower bounds of $(\tilde{x}_i^T \tilde{x}_j)$ in the objective function such that the objective function take its maximum value. For this goal, at first we write the objective function sentence by sentence as follows:

$$F(\alpha, \alpha') = \sum_{i=1}^{n} \alpha_i \tilde{y}_i - \sum_{i=1}^{n} \alpha_i^{'} \tilde{y}_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \tilde{x}_i^T \tilde{x}_j + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j^{'} \tilde{x}_i^T \tilde{x}_j$$

$$+ \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i^{'} \alpha_j \tilde{x}_i^T \tilde{x}_j - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i^{'} \alpha_j \tilde{x}_i^T \tilde{x}_j - \varepsilon \sum_{i=1}^{n} (\alpha_i + \alpha_i^{'})$$

Where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n, \alpha_1^{'}, \alpha_2^{'}, \ldots, \alpha_n^{'})$, $\alpha_i (i = 1, \ldots, n)$ corresponds to the slack variables ξ_i, and $\alpha_i^{'} (i = 1, \ldots, n)$ corresponds to the constraint whit slack $\xi_i^{'}$.

The terms with negative coefficient must take their lower bound and the terms whit positive coefficient must take their upper bound. Now if we show the minimum value of the terms $\tilde{x}_i^T \tilde{x}_j$ and $\tilde{y}_i$ by $(x_i^T x_j)^l$ and y_i^l, and their maximum value by $(x_i^T x_j)^u$ and y_i^u when $x_i \in \Re^m$ and $y_i \in \Re$ vary in intervals $[x_i^l, x_i^u]$ and $[y_i^l, y_i^u]$ respectively, to reach the maximal value for objective function, paying attention to the point $\alpha_i, \alpha_i^{'} \geq 0$ ($i \in \{1, 2, \ldots, n\}$), the problem (6) is converted to the following problem:

max $\quad F(\alpha, \alpha') =$

$$\sum_{i=1}^{n} \alpha_i y_i^u - \sum_{i=1}^{n} \alpha_i y_i^l - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j (x_i^T x_j)^l + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j^{'} (x_i^T x_j)^u$$

$$+ \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i^{'} \alpha_j (x_i^T x_j)^u - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i^{'} \alpha_j^{'} (x_i^T x_j)^l - \varepsilon \sum_{i=1}^{n} (\alpha_i + \alpha_i^{'})$$

s.t. $\quad \sum_{i=1}^{n} (\alpha_i - \alpha_i^{'}) = 0$ $\qquad\qquad\qquad\qquad$ (7)

$$0 \leq \alpha_i \leq C \qquad i = 1, \ldots, n$$

$$0 \leq \alpha_i^{'} \leq C \qquad i = 1, \ldots, n$$

Problem (7) is a conventional quadratic problem with constant objective coefficients and can be shown as following matrix form:

min $\quad \frac{1}{2} \alpha^T H \alpha + f^T \alpha$

With the previous problem constraints,

where $H = \begin{pmatrix} G^l & -G^u \\ -G^u & G^l \end{pmatrix}$, that G is a matrix with coefficients $G_{ij}^{\ l} = (x_i^T x_j)^l$ and

$G_{ij}^{\ u} = (x_i^T x_j)^u$ and $f = [\varepsilon - y_1^{\ u}, \varepsilon - y_2^{\ u}, ..., \varepsilon - y_n^{\ u}, \varepsilon + y_1^{\ l}, \varepsilon + y_2^{\ l}, ..., \varepsilon + y_n^{\ l}]$.

By the optimization conditions (5) we find the lower and upper bounds of the weight vector w as follows:

$$w^u = \sum_{i=1}^{n} \alpha_i x_i^{\ u} - \sum_{i=1}^{n} \alpha_i' x_i^{\ l} \tag{8}$$

$$w^l = \sum_{i=1}^{n} \alpha_i x_i^{\ l} - \sum_{i=1}^{n} \alpha_i' x_i^{\ u} \tag{9}$$

By definition the index sets $I_+ = \{i : \alpha_i > 0\}$ and $I_- = \{i : \alpha_i < 0\}$ and paying attention to this point that $\alpha_i \alpha_i' = 0$, so $I_+ \cap I_- = \phi$. So we can rewrite the equations (8) and (9) as follows:

$$w^u = \sum_{i \in I_+} \alpha_i x_i^{\ u} - \sum_{i \in I_-} \alpha_i' x_i^{\ l} \tag{10}$$

$$w^l = \sum_{i \in I_+} \alpha_i x_i^{\ l} - \sum_{i \in I_-} \alpha_i' x_i^{\ u} \tag{11}$$

Considering $\{x_i^{\ u}\}_{i \in I_+} \cup \{x_i^{\ l}\}_{i \in I_-}$ as a standard training sample set one can find the optimized bias b^u correspond to the weight vector w^u, and similarly by considering $\{x_i^{\ l}\}_{i \in I_+} \cup \{x_i^{\ u}\}_{i \in I_-}$ as a standard training sample set one can find the optimized bios b^l correspond to the weight vector w^l. Note that an arbitrary SV point, $(x_i^{\ SV}, y_i^{\ SV})$, satisfies in equivalence $y_i - w^T x_i - b = \varepsilon$, so substituting the above obtained values in (10) and (11), the related bias is achieved.

3. Experimental results

Some interval samples are generated randomly around the hyperplane $y = 3x + 4$, and in the figure 2 the center of samples are shown, (x_i, y_i), for all $i \in \{1, 2, ..., n\}$ we set: $x_i^{\ l} = x_i - 0.1$, $x_i^{\ u} = x_i + 0.1$, $y_i^{\ l} = y_i - 0.1$ and $y_i^{\ u} = y_i + 0.1$.

Solving the problem (6) the optimized weight vector w is obtained as the interval $[w^l, w^u] = [3.06, 3.24]$ the optimal bias $b^l = 4.05$ correspond to $w^l = 3.06$ and the optimal bias $b^u = 3.85$ correspond to $w^u = 3.24$. In the figure the hyperplane $(w^u)^T x + b^u = 3.24x + 3.85$ and the hyperplane $(w^l)^T x + b^l = 3.06x + 4.05$ are shown.
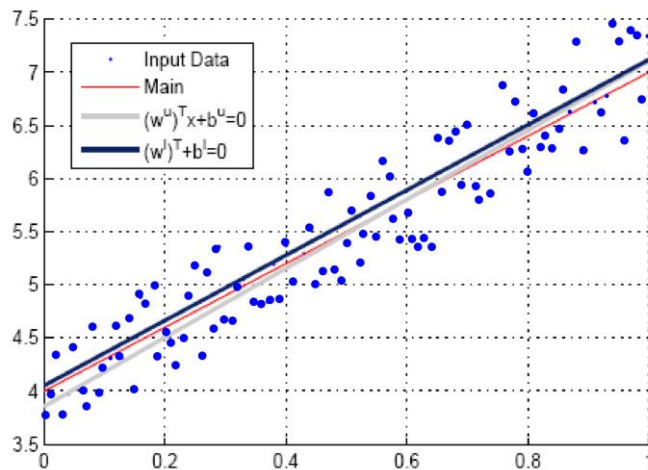
Figure 2.

4. Conclusion

In this work, we solved ISVR problem using SVM, obviously constant samples can be shown as intervals, so this version contains conventional SVR as well. Since in the real world, the parameters are seldom known and have to be estimated, almost, every problem with real parameters can be extended to a program with interval or fuzzy parameters. Clearly, we can extend ISVM to every kind of SVM, in the next work, we are going to solve interval support vector machine in One Class classification problem.

References

[1] S. T. Liu and R. T. Wang, "A numerical solution method to interval quadratic programming",   Applied mathematics And Computation 189 (2007) 1274-1281.

[2] V. Vapnik, The Nature of Statistical Learning Theory. New-York: Springer-Verlag, 1995.

[3] J.M. ŁęRski, An ε-insensitive approach to fuzzy clustering, Internat. J. Appl. Math. Comput. Sci. 11 (4) (2001) 993-1007

[5] J.M. ŁęRski, ε-insensitive fuzzy c-regression models: introduction to ε-insensitive fuzzy modeling, IEEE Trans. Systems Man Cybernet.-Part

B: Cybernet. 34 (1) (2004) 4-15.

[4] J.M. ŁęRski, Towards a robust fuzzy clustering, Fuzzy Sets and Systems 137 (2) (2003) 215-233.

[6] J.-T. Jeng, C. -C. Chuang, S.-F. Su, Support vector interval regression networks for interval regression analysis, Fuzzy Sets and Systems 138(2) (2003) 283-300.

[7] G. Castellano, A.M. Fanelli, C. Mencor, An empirical risk functional to improve learning in a neuro-fuzzy classifier, IEEE Trans. Systems, Man Cybernet.-Part B: Cybernet. 34 (1) (2004) 725-730.

[8] D. H. Hong, C. Hwang, Support vector fuzzy regression machines, Fuzzy Sets and Systems 138 (2) (2003) 271-281.

[9] Qi Wu, The forecasting model based on wavelet υ-support vector machine, Expert Systems with Applications 36 (2009) 7604-7610.