Research Article

Robust ε -Support Vector Regression

Yuan Lv and Zhong Gan

School of Mechanical Engineering, Northwestern Polytechnical University, Mailbox 301, No. 127 Youyi Road, Xi'an, Shaanxi 710072, China

Correspondence should be addressed to Yuan Lv; lv yuan85@live.com

Received 1 August 2013; Revised 25 November 2013; Accepted 25 November 2013; Published 20 February 2014

Academic Editor: Andrzej Swierniak

Copyright © 2014 Y. Lv and Z. Gan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Spheroid disturbance of input data brings great challenges to support vector regression; thus it is essential to study the robust regression model. This paper is dedicated to establish a robust regression model which makes the regression function robust against disturbance of data and system parameter. Firstly, two theorems have been given to show that the robust linear ε -support vector regression problem could be settled by solving the dual problems. Secondly, it has been focused on the development of robust support vector regression algorithm which is extended from linear domain to nonlinear domain. Finally, the numerical experiments result demonstrates the effectiveness of the models and algorithms proposed in this paper.

1. Introduction

As a new machine learning algorithm, support vector machine (SVM) is based on the statistical learning theory which was proposed by Vapnik [1] and Wolfe dual programming theory. Compared with other learning algorithms, SVM possesses a rigorous foundation of mathematical theory, well-generalized ability, and global optimum, with the result that is widely used in pattern recognition and functional regression.

In general, the input data is assumed to be accurate. However, in practice, the statistical errors and measuring errors of observed data are inevitable; that is, the data is disturbed. The effect of perturbation of input data on solving the regression problem is often neglected. Although SVR is robust, while the parameters are optimal, these parameters are usually reasonably sensitive to noise data and singular value. The robust optimization method, which focused on treatability of computation in the case of data points disturbing in convex sets, was first proposed by Soyster [2] and developed, respectively, by Ben-Tal and Nemirovski [3, 4] and El-Ghaoui et al. [5, 6]. Since then, Melvyn [7] presented more comprehensive collection of disturbance without increasing the computing difficulty, which enabled the rapid development of the robust optimization method. The research aimed on solving the programming problems with polyhedron disturbance or

spheroid disturbance. Bertsimas and Sim [8] and Alizadeh and Goldfarb [9] put forward the robust linear programming degenerated into a second-order cone programming, second-order cone programming into a semidefinite programming, and semi-definite programming into the NP problem. Zhixia et al. [10] applied this method to classification problem and proposed the robust support vector classification. This paper aims to study the regression problems of spheroid disturbance by means of the robust optimization method and to establish a robust regression model which makes the regression functions highly robust for perturbed data and system parameters.

2. Theory

2.1. Training Set. Assume that the training set is

$$T = \{(\chi_1, y_1), \dots, (\chi_m, y_m)\}, \quad (1)$$

where

$$\chi_i = \{x_i + r_i u_i \mid \|u_i\| \leq 1\}, \quad i = 1, \dots, m. \quad (2)$$

The input set x_i can be viewed as a sphere of radius r_i centering on the measured value x_i , r_i is closely related to measurement accuracy, and m is sample size.

2.2. Original Problem. Referring to the original problem of linear ε -support vector regression, it is not difficult to derive the original problem of robust linear ε -support vector regression, which is a convex quadratic programming, by replacing the original input with the input collection χ_i :

$$\begin{aligned} \min_{w,b,\xi,\bar{\xi}} \quad & \left(\frac{1}{2}\right) \cdot \|w\|^2 + C \sum_{i=1}^m (\xi_i + \bar{\xi}_i), \\ \text{s.t.} \quad & (w \cdot x_i + r_i u_i) + b - y_i \leq \varepsilon + \xi_i, \\ & y_i - (w \cdot x_i + r_i u_i) + b \leq \varepsilon + \bar{\xi}_i, \\ & \xi_i, \bar{\xi}_i \geq 0, \quad i = 1, \dots, m. \end{aligned} \quad (3)$$

Theorem 1. *The necessary and sufficient conditions for the solution of the original optimization problem (3) are that w^* , b^* , ξ^* , and $\bar{\xi}^*$ are the solutions for the following second-order cone programming:*

$$\begin{aligned} \min_{w,b,\xi,\bar{\xi},u,v,t} \quad & \left(\frac{1}{2}\right) \cdot (u - v) + C \sum_{i=1}^m (\xi_i + \bar{\xi}_i), \\ \text{s.t.} \quad & (w \cdot x_i) + b - y_i + r_i t \leq \varepsilon + \xi_i, \\ & y_i - ((w \cdot x_i) + b) + r_i t \leq \varepsilon + \bar{\xi}_i, \\ & \xi_i, \bar{\xi}_i \geq 0, \quad i = 1, \dots, m, \\ & u + v = 1, \\ & \begin{pmatrix} u \\ t \\ v \end{pmatrix} \in L^3, \\ & \begin{pmatrix} t \\ w \end{pmatrix} \in L^{n+1}, \end{aligned} \quad (4)$$

where n is the dimensionality of the input data χ_i .

Proof. Since

$$\begin{aligned} \max \{r_i (w \cdot u_i), \|u_i\| \leq 1\} &= r_i \|w\|, \\ \min \{r_i (w \cdot u_i), \|u_i\| \leq 1\} &= -r_i \|w\|. \end{aligned} \quad (5)$$

The original problem (3) can be converted to

$$\begin{aligned} \min_{w,b,\xi,\bar{\xi}} \quad & \left(\frac{1}{2}\right) \cdot \|w\|^2 + C \sum_{i=1}^m (\xi_i + \bar{\xi}_i), \\ \text{s.t.} \quad & (w \cdot x_i) + b - y_i + r_i \|w\| \leq \varepsilon + \xi_i, \\ & y_i - ((w \cdot x_i) + b) + r_i \|w\| \leq \varepsilon + \bar{\xi}_i, \\ & \xi_i, \bar{\xi}_i \geq 0, \quad i = 1, \dots, m. \end{aligned} \quad (6)$$

If variable t is introduced, satisfying $\|w\| \leq t$, the above problem could prove to be a second-order cone programming:

$$\begin{aligned} \min_{w,b,\xi,\bar{\xi}} \quad & \left(\frac{1}{2}\right) \cdot t^2 + C \sum_{i=1}^m (\xi_i + \bar{\xi}_i), \\ \text{s.t.} \quad & (w \cdot x_i) + b - y_i + r_i t \leq \varepsilon + \xi_i, \\ & y_i - ((w \cdot x_i) + b) + r_i t \leq \varepsilon + \bar{\xi}_i, \\ & \xi_i, \bar{\xi}_i \geq 0, \quad i = 1, \dots, m, \\ & \|w\| \leq t. \end{aligned} \quad (7)$$

When variables u and v are brought in, u and v are supposed to satisfy the linear constraint $u + v = 1$ and the second-order cone constraint $(t^2 + v^2)^{1/2} \leq u$, and the nonlinear terms t^2 in the objective function are replaced with $u - v$; therefore, the problem (7) can be written as the problem (4). From the above, the problem (3) is equivalent to the problem (4). $\square$

2.3. Dual Problem. Introduce the Lagrange function:

$$\begin{aligned} L = & \left(\frac{1}{2}\right) \cdot (u - v) + C \sum_{i=1}^m (\xi_i + \bar{\xi}_i) \\ & + \sum_{i=1}^m \alpha_i ((w \cdot x_i) + b - y_i + r_i t - \varepsilon - \xi_i) \\ & + \sum_{i=1}^m \beta_i (y_i - ((w \cdot x_i) + b) + r_i t - \varepsilon - \bar{\xi}_i) \\ & - \sum_{i=1}^m \theta_i \xi_i - \sum_{i=1}^m \eta_i \bar{\xi}_i + \varphi \cdot (u + v - 1) \\ & - z_u u - z_v v - \delta t - z_t t - z_w^T w, \end{aligned} \quad (8)$$

where $\alpha, \beta, \theta, \eta \in R^m$, $\varphi, z_u, z_v, \delta, z_t \in R$, $z_w \in R^n$ are the multipliers vectors corresponding to the constraints. Eliminating variables θ, η, z_t , and z_w , the following theorem can be obtained.

Theorem 2. *The following second-order cone programming is the dual problem of the second-order cone programming (4).*

$$\begin{aligned} \max_{\alpha,\beta,\varphi,\delta,z_u,z_v} \quad & - \sum_{i=1}^m \alpha_i (y_i + \varepsilon) + \sum_{i=1}^m \beta_i (y_i - \varepsilon) - \varphi, \\ \text{s.t.} \quad & z_u - \varphi = \frac{1}{2}, \\ & z_v - \varphi = -\frac{1}{2}, \\ & 0 \leq \alpha_i \leq C, \\ & 0 \leq \beta_i \leq C, \\ & z_u \geq (\delta^2 + z_v^2)^{1/2}, \end{aligned}$$

$$\begin{aligned}
& \sum_{i=1}^m \alpha_i - \sum_{i=1}^m \beta_i = 0, \\
& i = 1, \dots, m, \quad j = 1, \dots, m, \\
& \delta \leq \sum_{i=1}^m r_i (\alpha_i + \beta_i) \\
& - \left(\sum_{i=1}^m \sum_{j=1}^m (\alpha_i - \beta_i)(\alpha_j - \beta_j)(x_i \cdot x_j) \right)^{1/2}.
\end{aligned} \tag{9}$$

Proof. The feasible region of the second-order cone programming problem (4) is

$$\begin{aligned}
D = \left\{ w, b, \xi, \bar{\xi}, u, v, t \mid (w \cdot x_i) + b - y_i \right. \\
+ r_i t \leq \varepsilon + \xi_i, y_i - ((w \cdot x_i) + b) \\
+ r_i t \leq \varepsilon + \bar{\xi}_i, \xi_i, \bar{\xi}_i \geq 0, \\
i = 1, \dots, m, u \geq (t^2 + v^2)^{1/2}, \\
\left. t \geq \left(\sum_{j=1}^n w_j^2 \right)^{1/2}, j = 1, \dots, n; u + v = 1 \right\}.
\end{aligned} \tag{10}$$

The feasible region of the dual problem (9) is

$$\begin{aligned}
E = \left\{ \alpha, \beta, \varphi, \delta, z_u, z_v \mid 0 \leq \alpha_i \leq C, \right. \\
0 \leq \beta_i \leq C, z_u \geq (\delta^2 + z_v^2)^{1/2}, \\
\delta \leq \sum_{i=1}^m r_i \cdot (\alpha_i + \beta_i) \\
- \left(\sum_{i=1}^m \sum_{j=1}^m (\alpha_i - \beta_i)(\alpha_j - \beta_j)(x_i \cdot x_j) \right)^{1/2}, \\
z_u - \varphi = \frac{1}{2}, z_v - \varphi = -\frac{1}{2}, \\
\left. \sum_{i=1}^m \alpha_i - \sum_{i=1}^m \beta_i = 0, i = 1, \dots, m, j = 1, \dots, m \right\}.
\end{aligned} \tag{11}$$

Assuming that the optimal value of the problem (4) is p^* ,

$$\begin{aligned}
& p^* \\
& = \inf \left\{ \left(\frac{1}{2} \right) \cdot (u - v) + C \sum_{i=1}^m (\xi_i + \bar{\xi}_i) \mid w, b, \xi, \bar{\xi}, u, v, t \in D \right\}.
\end{aligned} \tag{12}$$

Proceeding from the lower bound of the estimate of p^* , the Lagrange function of the second-order cone programming problem (4) is introduced and written as $L(w, b, \xi, \bar{\xi}, u, v, t, \alpha, \beta, \varphi, \delta, z_u, z_v)$:

$$\begin{aligned}
& \forall w, b, \xi, \bar{\xi}, u, v, t \in D, \alpha, \beta, \varphi, \delta, z_u, z_v \in E, \\
& L(w, b, \xi, \bar{\xi}, u, v, t, \alpha, \beta, \varphi, \delta, z_u, z_v) \\
& \leq \left(\frac{1}{2} \right) \cdot (u - v) + C \sum_{i=1}^m (\xi_i + \bar{\xi}_i), \\
& \inf_{w, b, \xi, \bar{\xi}, u, v, t \in R^n} L(w, b, \xi, \bar{\xi}, u, v, t, \alpha, \beta, \varphi, \delta, z_u, z_v) \\
& \leq \inf_{w, b, \xi, \bar{\xi}, u, v, t \in D} L(w, b, \xi, \bar{\xi}, u, v, t, \alpha, \beta, \varphi, \delta, z_u, z_v) \\
& \leq \inf_{w, b, \xi, \bar{\xi}, u, v, t \in D} \left(\left(\frac{1}{2} \right) \cdot (u - v) + C \sum_{i=1}^m (\xi_i + \bar{\xi}_i) \right).
\end{aligned} \tag{13}$$

let

$$\begin{aligned}
& g(\alpha, \beta, \varphi, \delta, z_u, z_v) \\
& = \inf_{w, b, \xi, \bar{\xi}, u, v, t \in R^n} L(w, b, \xi, \bar{\xi}, u, v, t, \alpha, \beta, \varphi, \delta, z_u, z_v);
\end{aligned} \tag{14}$$

thus

$$g(\alpha, \beta, \varphi, \delta, z_u, z_v) \leq p^*. \tag{15}$$

Formula (15) implies $\forall w, b, \xi, \bar{\xi}, u, v, t \in D, \alpha, \beta, \varphi, \delta, z_u, z_v \in E$, $g(\alpha, \beta, \varphi, \delta, z_u, z_v)$ is the lower bound of p^* .

To get the optimal value, the maximal value of $g(\alpha, \beta, \varphi, \delta, z_u, z_v)$, $\forall w, b, \xi, \bar{\xi}, u, v, t \in D, \alpha, \beta, \varphi, \delta, z_u, z_v \in E$ is need to be obtained. So, the dual problem (9) is deduced. $\square$

2.4. Robust SVR Algorithm

2.4.1. Linear Robust SVR Algorithm. (1) The given training data should be written as the way of formulae (1)~(2).

(2) By constructing and solving the second-order cone programming problem (9), the solution $(\alpha_i^*, \beta_i^*, \varphi^*, \delta^*, z_u^*, z_v^*)$ can be obtained. Besides, it is necessary to select an appropriate penalty parameter $C > 0$.

(3) The solution (w^*, b^*) of the original problem (4) is acquired in the following way:

$$w^* = \frac{\delta^* \sum_{i=1}^m (\alpha_i^* - \beta_i^*) x_i}{(\sum_{i=1}^m (\alpha_i^* + \beta_i^*) r_i - \delta^*)}. \tag{16}$$

Select a component α_j of α or a component β_k of β , which is located in the open interval $(0, C)$; if α_j is chosen, it is got as follows:

$$b^* = \frac{y_j - \delta^* \sum_{i=1}^m (\alpha_i^* - \beta_i^*) (x_i \cdot x_j)}{(\sum_{i=1}^m (\alpha_i^* + \beta_i^*) r_i - \delta^*) + \varepsilon + r_i \delta^*}. \tag{17}$$

If β_k is chosen, it is got as follows:

$$b^* = \frac{y_k - \delta^* \sum_{i=1}^m (\alpha_i^* - \beta_i^*) (x_i \cdot x_k)}{(\sum_{i=1}^m (\alpha_i^* + \beta_i^*) r_i - \delta^*) - \varepsilon - r_i \delta^*}. \quad (18)$$

(4) Then regression function is established:

$$f(x) = \frac{\delta^* \sum_{i=1}^m (\alpha_i^* - \beta_i^*) (x_i \cdot x)}{(\sum_{i=1}^m (\alpha_i^* + \beta_i^*) r_i - \delta^*) + b^*}. \quad (19)$$

2.4.2. Nonlinear Robust SVR Algorithm. By making the following modifications to the algorithm shown in Section 2.4.1, the nonlinear robust SVR algorithm can be constructed:

- (1) Kernel function $K(x_i, x_j)$ (common kernel functions include Gaussian kernel function, polynomial kernel function, Cauchy kernel function, and Laplace kernel function) is introduced; replace the inner products $(x_i \cdot x_j)$ and $(x_i \cdot x)$ appearing in the linear robust SVR algorithm shown in Section 2.4.1, with $K(x_i, x_j)$ and $K(x_i, x)$;
- (2) Substitute R_i for r_i appearing in the second-order cone programming (9).

To get R_i , the mapping should be written as $X = \Phi(x)$; obviously, $\|x_i - x_j\| \leq r_i$ is equivalent to

$$\begin{aligned} & \|\Phi(x_i) - \Phi(x_j)\|^2 \\ &= ((\Phi(x_i) - \Phi(x_j)) \cdot (\Phi(x_i) - \Phi(x_j))) \\ &= K(x_i, x_i) - 2K(x_i, x_j) + K(x_j, x_j) \\ &= F(\|x_i - x_j\|) \leq R_i^2. \end{aligned} \quad (20)$$

So $R_i = F(r_i)$.

3. Numerical Experiments

3.1. Linear Numerical Experiment

Example 1. Given that objective function is $y = 2x$ ($x \in [1, 10]$), the training set could be written as the way of formulae (1)~(2), where x_i is 10 equidistant points distributed in the interval $[1, 10]$. r_i can be regarded as a constant which equals 0.5. The disturbance of data u_i is a random number which is located in the interval $[-1, 1]$, where $i = 10$. The methods of standard SVR and robust SVR shown in Section 2.4.1 are employed to regress and fit the objective function. Then the errors of ten test points, which are randomly selected, are analysed and compared. The results are shown in Table 1.

3.2. Nonlinear Numerical Experiments

Example 2. Given that objective function is $y = x^2$ ($x \in [0.5, 10]$), the training set could be written in the way of formulae (1)~(2), where x_i is 20 equidistant points distributed in the interval $[0.5, 10]$. r_i can be regarded as a constant which

TABLE 1: Table of error analysis of linear function regression.

Test points	Algorithm	
	Linear standard SVR	Linear robust SVR
Point 1	0.4388	0.0062
Point 2	0.3316	0.0065
Point 3	0.2244	0.0068
Point 4	0.1172	0.0072
Point 5	0.0100	0.0075
Point 6	0.0972	0.0078
Point 7	0.2044	0.0082
Point 8	0.3117	0.0085
Point 9	0.4189	0.0089
Point 10	0.5261	0.0092
AVG	0.2680	0.0077

TABLE 2: Table of error analysis of nonlinear function regression.

Test points	Algorithm	
	Nonlinear standard SVR	Nonlinear robust SVR
Point 1	0.1898	0.0511
Point 2	0.0668	0.0466
Point 3	0.0788	0.0385
Point 4	0.2183	0.0270
Point 5	0.2929	0.0119
Point 6	0.3589	0.0066
Point 7	0.5538	0.0287
Point 8	0.0887	0.0542
Point 9	0.0331	0.0833
Point 10	0.1305	0.1158
AVG	0.2012	0.0464

equals 0.05. The disturbance of data u_i is a random number which is located in the interval $[-1, 1]$, where $i = 20$. The methods of standard SVR and robust SVR shown in Section 2.4.2 are employed to regress and fit the objective function. Then the errors of ten test points, which are randomly selected, are analysed and compared. The results are given in Table 2.

3.3. Discussion. Obviously, the system parameters are very robust. In addition, it is shown that the forecast errors of the robust method are lower than that of the standard SVM method in Table 1. A similar conclusion can also be presented in Table 2. In short, the robust SVR method has been validated to be of advantage in the field of regression analysis with perturbation problem.

4. Conclusions

- (1) In the paper, a robust ε -SVR model has been proposed, considering the input data as the convex set of spheroid disturbance, and two important theorems are also given: it is proved by Theorem 1 that the

convex quadratic programming problem can be written as the second-order cone programming. Based on duality theory, the dual problem can be deduced by Theorem 2. By solving the dual problem, the optimal solution of original problem is obtained.

- (2) By virtue of kernel function, the robust linear ε -SVR model could be promoted to the nonlinear domains. Finally, the numerical experiments demonstrate the robustness and effectiveness of the proposed algorithm.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The theorems and algorithms proposed in this paper are the extension of robust support vector classification method [10], which provides a great assistance for the authors. The authors are here to express their sincere gratitude.

References

- [1] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 1999.
- [2] A. L. Soyster, "Inexact linear programming with generalized resource sets," *European Journal of Operational Research*, vol. 3, no. 4, pp. 316–321, 1979.
- [3] A. Ben-Tal and A. Nemirovski, "Robust convex optimization," *Mathematics of Operations Research*, vol. 23, no. 4, pp. 769–805, 1998.
- [4] A. Ben-Tal and A. Nemirovski, "Robust solutions of uncertain linear programs," *Operations Research Letters*, vol. 25, no. 1, pp. 1–13, 1999.
- [5] L. El Ghaoui and H. Lebret, "Robust solutions to least-squares problems with uncertain data," *SIAM Journal on Matrix Analysis and Applications*, vol. 18, no. 4, pp. 1035–1064, 1997.
- [6] L. El Ghaoui, F. Oustry, and H. Lebret, "Robust solutions to uncertain semidefinite programs," *SIAM Journal on Optimization*, vol. 9, no. 1, pp. 33–52, 1999.
- [7] S. Melvyn, *Robust optimization [Ph.D. thesis]*, 2004.
- [8] D. Bertsimas and M. Sim, "Tractable approximations to robust conic optimization problems," *Mathematical Programming*, vol. 107, no. 1-2, pp. 5–36, 2006.
- [9] F. Alizadeh and D. Goldfarb, "Second-order cone programming," *Mathematical Programming, Series B*, vol. 95, no. 1, pp. 3–51, 2003.
- [10] Y. Zhixia, T. Yingjie, and D. Naiyang, "Second order cone programming formulations for robust support vector ordinal regression machine," in *Proceedings of the 1st International Symposium on Optimization and Systems Biology*, pp. 332–340, 2013.