Knowledge-Based Systems 33 (2012) 92–101

Contents lists available at SciVerse ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

A weighted twin support vector regression

Yitian Xu *, Laisheng Wang

College of Science, China Agricultural University, Beijing 100083, China

ARTICLE INFO

Article history:
Received 24 June 2011
Received in revised form 10 March 2012
Accepted 11 March 2012
Available online 21 March 2012

Keywords:
SVR
TSVR
Up- and down-bound functions
Weighted coefficient
Weighted TSVR

ABSTRACT

Twin support vector regression (TSVR) is a new regression algorithm, which aims at finding ϵ-insensitive up- and down-bound functions for the training points. In order to do so, one needs to resolve a pair of smaller-sized quadratic programming problems (QPPs) rather than a single large one in a classical SVR. However, the same penalties are given to the samples in TSVR. In fact, samples in the different positions have different effects on the bound function. Then, we propose a weighted TSVR in this paper, where samples in the different positions are proposed to give different penalties. The final regressor can avoid the over-fitting problem to a certain extent and yield great generalization ability. Numerical experiments on one artificial dataset and nine benchmark datasets demonstrate the feasibility and validity of our proposed algorithm.

1. Introduction

Support vector machine (SVM), motivated by the Vapnik–Chervonenkis (VC) dimensional theory and the statistical learning theory [15], is a promising technique. Many papers exploiting it made the state of the art and one of the most used classifiers. Compared with other machine learning approaches like artificial neural networks [14], SVM has many advantages. First, SVM solves a QPP, assuring that once an optimal solution is obtained, it is the unique (global) solution. Second, SVM derives its sparse and robust solution by maximizing the margin between the two classes. Third, SVM implements the structural risk minimization principle rather than the empirical risk minimization principle, which minimizes the upper bound of the generalization error. SVM has been successfully applied in various aspects ranging from remote sensing image classification [11], text classification [18] to business prediction [10].

However, one of the main challenges for the standard SVM is the high computational complexity. The computational complexity of the SVM is n^3, where n is the total size of training data. In order to improve the computational speed of SVM, Jayadeva et al. [5] proposed a twin support vector machine (TSVM) for binary data classification in the spirit of the proximal SVM [2,4,3]. TSVM generates two nonparallel hyper-planes by solving two smaller-sized QPPs such that each hyper-plane is closer to one class and as far as possible from the other. The strategy of solving two smaller-sized QPPs, rather than a single large one, makes the learning speed of TSVM approximately four times faster than that of the standard SVM. At present, TSVM has become one of the popular methods because of its low computational complexity. Many variants of TSVM have been proposed by Peng [12], Kumar and Gopal [7], Jayadeva et al. [6], Khemchandani et al. [9]. Certainly, the above algorithms are suitable to the classification problems. As for the regression problem, Peng [13] proposed an efficient TSVR.

In TSVR, the same penalties are given to the samples. However, as samples locate in the different positions, it is more reasonable to give different penalties to them. Inspired by the above studies, we introduce two weighted coefficients σ_1 and σ_2 [1,19,16] into the TSVR and propose a weighted TSVR in this paper. By dividing the whole plane into different parts, we bring different penalties to the samples depending on their different positions.

The effectiveness of our proposed algorithm is demonstrated by numerical experiments on one artificial dataset and nine benchmark datasets. In the artificial experiment, we preliminarily determine the range of the penalty parameter σ, and it is helpful to the choice of the parameter σ in the following benchmark experiments. While we investigate the distributions of samples. The experimental results on nine benchmark datasets show that the weighted TSVR achieves significant performance in comparison with SVR and TSVR.

The paper is organized as follows. Section 2 outlines the SVR and TSVR. A weighted TSVR is proposed in Section 3, which includes both the linear and nonlinear cases. Section 4 performs experiments on one artificial dataset and nine benchmark datasets to investigate the effectiveness of the weighted TSVR. The last section concludes the conclusions.

* Corresponding author. Tel.: +86 10 62737077; fax: +86 10 62736777.
E-mail address: xytshuxue@126.com (Y. Xu).

0950-7051/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.knosys.2012.03.013

2. SVR and TSVR

In this section, we give a brief description of the SVR and TSVR. Given a training set $T=\{(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)\}$, where $x_i \in R^d$ and $y_i \in R$. For the sake of conciseness, let matrix $A=(x_1,x_2,\ldots,x_n)$ and matrix $Y=(y_1,y_2,\ldots,y_n)$.

2.1. Support vector regression

The nonlinear SVR seeks to find a regression function $f(x)=w^T\phi(x)+b$ in a high dimensional feature space tolerating the small error in fitting the given dataset. This can be achieved by utilizing the ϵ-insensitive loss function that sets an ϵ-insensitive "tube" around the data, within which errors are discard. The nonlinear SVR can be obtained by resolving the following QPP:

$$\begin{aligned} \min_{w,b,\xi,\xi^*} \quad & \frac{1}{2}\|w\|^2 + c\cdot(e^T\xi + e^T\xi^*), \\ \text{s.t.} \quad & (\phi^T(A)w+eb) - Y \leqslant e\epsilon + \xi, \\ & Y - (\phi^T(A)w+eb) \leqslant e\epsilon + \xi^*, \\ & \xi \geqslant 0e, \quad \xi^* \geqslant 0e, \end{aligned} \tag{1}$$

where c is a parameter chosen a priori, which weights the tradeoff between the fitting errors and flatness of the regression function, ξ and ξ^* are the slack vectors reflecting whether the samples locate into the ϵ-tube or not, e is the vector of ones of appropriate dimensions.

By introducing the Lagrangian multiplies α and α^*, we can derive the dual problem of the QPP (1) as follows:

$$\begin{aligned} \max_{\alpha,\alpha^*\in R^n} \quad & -\frac{1}{2}(\alpha^*-\alpha)^T K(A,A^T)(\alpha^*-\alpha) + Y^T(\alpha^*-\alpha) + \epsilon e^T(\alpha^*+\alpha), \\ \text{s.t.} \quad & e^T(\alpha^*-\alpha) = 0, \\ & 0e \leqslant \alpha, \alpha^* \leqslant ce. \end{aligned} \tag{2}$$

Once the QPP (2) is resolved, we can achieve its solution $\alpha^{(*)} = (\alpha_1,\alpha_1^*,\alpha_2,\alpha_2^*,\ldots,\alpha_n,\alpha_n^*)$ and threshold b, and then obtain the regression function,

$$f(x) = \sum_{i=1}^n (\alpha_i^* - \alpha_i)K(x_i,x) + b. \tag{3}$$

Here, $K(x_i,x) = (\phi(x_i)\cdot\phi(x))$ represents a kernel function which gives the dot product in the high dimensional feature space. α and α^* are lagrange multipliers that satisfy $\alpha_i\alpha_i^* = 0$, $i=1,2,\ldots,n$. We can find that the regressor $f(x)$ is only decided by the samples (support vectors) whose lagrange multipliers $\alpha_i \neq 0$ or $\alpha_i^* \neq 0$. Moreover, lagrange multipliers α_i, $\alpha_i^* = 0$ for most samples. Therefore SVR owns sparsity.

2.2. Twin support vector regression

In order to improve the computational speed, Peng [13] proposed an efficient TSVM for the regression problem, termed as TSVR. TSVR generates an ϵ-insensitive down-bound function $f_1(x) = w_1^Tx + b_1$ and an ϵ-insensitive up-bound function $f_2(x) = w_2^Tx + b_2$. TSVR is illustrated in Fig. 1.

The final regressor $f(x)$ is decided by the mean of these two bound functions, i.e.,

$$f(x) = \frac{1}{2}(f_1(x)+f_2(x)) = \frac{1}{2}(w_1+w_2)^Tx + \frac{1}{2}(b_1+b_2). \tag{4}$$

TSVR is obtained by solving the following pair of QPPs,

$$\begin{aligned} \min_{w_1,b_1,\xi} \quad & \frac{1}{2}\|Y - e\epsilon_1 - (Aw_1+eb_1)\|^2 + c_1e^T\xi, \\ \text{s.t.} \quad & Y - (Aw_1+eb_1) \geqslant e\epsilon_1 - \xi, \\ & \xi \geqslant 0e, \end{aligned} \tag{5}$$

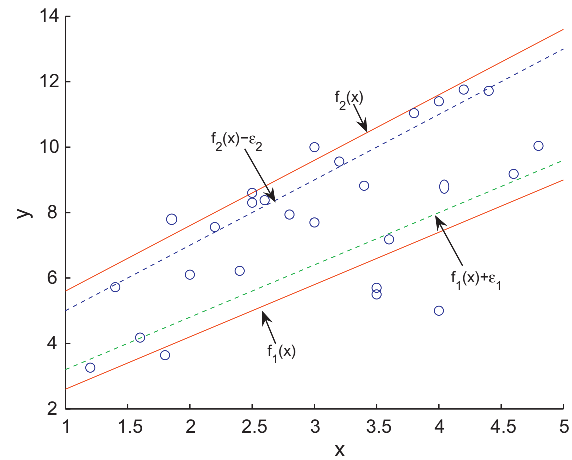

Fig. 1. Illustration of the TSVR.

and

$$\begin{aligned} \min_{w_2,b_2,\eta} \quad & \frac{1}{2}\|Y + e\epsilon_2 - (Aw_2+eb_2)\|^2 + c_2e^T\eta, \\ \text{s.t.} \quad & (Aw_2+eb_2) - Y \geqslant e\epsilon_2 - \eta, \\ & \eta \geqslant 0e, \end{aligned} \tag{6}$$

where c_1, c_2, ϵ_1 and ϵ_2 are parameters chosen a priori, ξ and η are slack vectors. By introducing the Lagrangian multipliers α and β, we can derive their dual problems as follows,

$$\begin{aligned} \max_{\alpha} \quad & -\frac{1}{2}\alpha^TG(G^TG)^{-1}G^T\alpha + f^TG(G^TG)^{-1}G^T\alpha - f^T\alpha, \\ \text{s.t.} \quad & 0e \leqslant \alpha \leqslant c_1e, \end{aligned} \tag{7}$$

and

$$\begin{aligned} \max_{\beta} \quad & -\frac{1}{2}\beta^TG(G^TG)^{-1}G^T\beta - h^TG(G^TG)^{-1}G^T\beta + h^T\beta, \\ \text{s.t.} \quad & 0e \leqslant \beta \leqslant c_2e, \end{aligned} \tag{8}$$

where $G=[A\ e]$, $f=Y-e\epsilon_1$, and $h=Y+e\epsilon_2$.

Once the dual problems (7) and (8) are solved, we can get $[w_1\ b_1]$ and $[w_2\ b_2]$ in (4) as follows:

$$[w_1\ b_1]^T = (G^TG)^{-1}G^T(f-\alpha), \tag{9}$$

$$[w_2\ b_2]^T = (G^TG)^{-1}G^T(h+\beta). \tag{10}$$

For the nonlinear case, TSVR resolves the following pair of QPPs:

$$\begin{aligned} \min_{w_1,b_1,\xi} \quad & \frac{1}{2}\|Y - e\epsilon_1 - (K(A,A^T)w_1+eb_1)\|^2 + c_1e^T\xi, \\ \text{s.t.} \quad & Y - (K(A,A^T)w_1+eb_1) \geqslant e\epsilon_1 - \xi, \\ & \xi \geqslant 0e, \end{aligned} \tag{11}$$

and

$$\begin{aligned} \min_{w_2,b_2,\eta} \quad & \frac{1}{2}\|Y + e\epsilon_2 - (K(A,A^T)w_2+eb_2)\|^2 + c_2e^T\eta, \\ \text{s.t.} \quad & K(A,A^T)w_2+eb_2 - Y \geqslant e\epsilon_2 - \eta, \\ & \eta \geqslant 0e. \end{aligned} \tag{12}$$

Similarly, we can derive the dual problems of the QPPs (11) and (12) as follows:

$$\begin{aligned} \max_{\alpha} \quad & -\frac{1}{2}\alpha^TH(H^TH)^{-1}H^T\alpha + f^TH(H^TH)^{-1}H^T\alpha - f^T\alpha, \\ \text{s.t.} \quad & 0e \leqslant \alpha \leqslant c_1e, \end{aligned} \tag{13}$$