Neurocomputing 171 (2016) 649–663

Contents lists available at ScienceDirect

Neurocomputing

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

A novel parametric-insensitive nonparallel support vector machine for regression

Zhi-Min Yang [a,b], Xiang-Yu Hua [a,*], Yuan-Hai Shao [b,*], Ya-Fen Ye [b,c]

[a] School of Economics and Management, Zhejiang University of Technology, Hangzhou 310024, PR China
[b] Zhijiang College, Zhejiang University of Technology, Hangzhou 310024, PR China
[c] School of Economics, Zhejiang University, Hangzhou 310024, PR China

ARTICLE INFO

Article history:
Received 15 September 2014
Received in revised form
2 July 2015
Accepted 2 July 2015
Communicated by W.S. Hong
Available online 11 July 2015

Keywords:
Support vector machine
Regression estimation
Bound estimation
Noise heteroscedastic model
Parametric insensitive model

ABSTRACT

In this paper, a novel parametric-insensitive nonparallel support vector regression (PINSVR) algorithm for data regression is proposed. PINSVR indirectly finds a pair of nonparallel proximal functions with a pair of different parametric-insensitive nonparallel proximal functions by solving two smaller sized quadratic programming problems (QPPs). By using new parametric-insensitive loss functions, the proposed PINSVR automatically adjusts a flexible parametric-insensitive zone of arbitrary shape and minimal size to include the given data to capture data structure and boundary information more accurately. The experiment results compared with the ε-SVR, ε-TSVR, and TPISVR indicate that our PINSVR not only obtains comparable regression performance, but also obtains better bound estimations.

1. Introduction

Powerful tools for data classification and regression in machine learning, support vector machines (SVMs) [1–5], including support vector classification (SVC) and support vector regression (SVR), have been successfully applied to a variety of real-world problems [6–12] in the past decades. The structural risk minimization principle is implemented in SVMs by minimizing the upper bound of the generalization error [13,14]. For SVR, i.e., ε-support vector regression (ε-SVR) [1–4], it finds a function $f(x)$ such that, on one hand, more training samples locate in the ε-intensive tube between $f(x)-\varepsilon$ and $f(x)+\varepsilon$, on the other hand, the function $f(x)$ is as flat as possible. ε-SVR does not care about errors as long as the samples are inside the ε-intensive tube, only those outside the ε-intensive tube are punished.

Recently, in the spirit of twin support vector machine (TWSVM) [15], some novel SVR algorithms for data regression, including ε-twin SVR (ε-TSVR) [16] and twin parametric insensitive SVR (TPISVR) [17], have been proposed. These algorithms determine indirectly the regressor through a pair of nonparallel proximal functions solved by two smaller sized QPPs instead of the larger single 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. However, both ε-SVR and ε-TSVR assume that the noise level on training data is uniform throughout the domain, or at least, its functional dependency is known beforehand [17–20]. The assumption of a uniform noise model is not always satisfied in the real-world. For instance, for the heteroscedastic noise structure, that is, the noise strongly depends on the inputs. For TPISVR, it determines the parametric insensitive down- and up-bound functions by using the parametric-insensitive loss function, so it is more suitable for the case that the noise is heteroscedastic. However, TPISVR only aims at minimizing the empirical risk, but not embeds any structural risk of data into the learning process, which leads the down- and up-bound functions to be possibly contaminated by noise samples.

In this paper, we propose a novel SVR model for data regression, termed the parametric-insensitive nonparallel support vector regression (PINSVR). Our PINSVR indirectly finds a pair of nonparallel proximal functions with a pair of different parametric-insensitive nonparallel proximal functions by solving two smaller sized QPPs, which leads it to be more suitable for the heteroscedastic noise structure. By introducing the parametric-insensitive loss functions, our PINSVR automatically adjusts a flexible parametric-insensitive zone of arbitrary shape and minimal size to include the given data to capture data structure and

* Corresponding authors. Tel.: +86 0571 87313679; fax: +86 0571 87313551.
E-mail addresses: yzm9966@126.com (Z.-M. Yang), hxy9069@163.com (X.-Y. Hua), shaoyuanhai21@163.com (Y.-H. Shao), yafenye@163.com (Y.-F. Ye).

http://dx.doi.org/10.1016/j.neucom.2015.07.003
0925-2312/

boundary information more accurately. Compared with the ε-SVR, ε-TSVR, and TPISVR, the merits of our PINSVR are as follows: (i) the loss functions used in our PINSVR are different from those in ε-SVR, ε-TSVR, and TPISVR, and the unique loss setting makes our PINSVR more adaptive; (ii) only the empirical risk is minimized in TPISVR, whereas in our PINSVR the structural risk is minimized by adding a regularization term with the idea of maximizing some margin. Computational comparisons with ε-SVR, ε-TSVR, and TPISVR in terms of generalization performance have been made on several artificial, benchmark and real practical datasets, indicating our PINSVR not only obtains comparable regression performance, but also obtains better bound estimations.

The rest of this paper is organized as follows: Section 2 briefly introduces classical ε-SVR, ε-TSVR, and TPISVR. Section 3 presents the proposed parametric-insensitive nonparallel support vector regression (PINSVR) model. Experimental results on several artificial, benchmark and real practical datasets are given in Section 4. Some conclusions and remarks are drawn in Section 5.

2. Backgrounds

Consider the following regression problem, suppose that the training set is denoted by $(A,Y)=\{(x_1,y_1),\dots,(x_m,y_m)\}\subset R^n\times R$, where A is a $m\times n$ matrix and the i-th row $A_i\in R^n$ represents the i-th training sample, $i=1,2,\dots,m$. Let $Y=(y_1,\dots,y_m)$ denote the response vector of training samples, where $y_i\in R$. Here, some methods that are closely related to our method are briefly described, including ε-SVR [1], ε-TSVR [16], and TPISVR [17]. For simplicity, only the linear regressors are considered.

2.1. ε-Support vector regression (ε-SVR)

The classical ε-SVR searches for an optimal linear regression function

$$f(x)=w^Tx+b \tag{1}$$

where $w\in R^n$ and $b\in R$. To measure the empirical risk, the ε-insensitive loss function [1,2,7]

$$L^{\varepsilon}(x,y,f)=|y-f(x)|_{\varepsilon}=\begin{cases}0 & \text{if}\,|y-f(x)|\le\varepsilon\\ |y-f(x)|-\varepsilon & \text{others}\end{cases} \tag{2}$$

is considered that sets a nonnegative ε tube around the data, within which errors are discarded. By introducing the regularization term $\frac{1}{2}\|w\|^2$ and the slack variables ξ and η, the primal problem of ε-SVR can be expressed as

$$\begin{aligned}\min_{w,b,\xi,\eta}\quad & \frac{1}{2}\|w\|^2+c(e^T\xi+e^T\eta)\\ \text{s.t.}\quad & Y-(Aw+eb)\ge-\varepsilon e-\xi,\quad \xi\ge0\\ & (Aw+eb)-Y\ge-\varepsilon e-\eta,\quad \eta\ge0\end{aligned} \tag{3}$$

where c is a positive parameter determining the trade-off between the empirical risk and the regularization term. Note that a small $\|w\|^2$ corresponds to the linear function (1) that is flat [7]. In the case of SVC, the structural risk minimization principle is implemented by this regularization term $\frac{1}{2}\|w\|^2$. In the case of SVR, this term is also added to minimize the structural risk. An intuitive two-dimensional geometric interpretation and loss setting for ε-SVR are shown in Fig. 1.

By introducing the Lagrangian multiplier technique, we obtain the following dual QPP for (3):

$$\begin{aligned}\min_{\alpha,\beta}\quad & \varepsilon e^T(\alpha+\beta)-Y^T(\alpha-\beta)+\frac{1}{2}(\alpha-\beta)^TA^TA(\alpha-\beta)\\ \text{s.t.}\quad & e^T(\alpha-\beta)=0,\quad 0\le\alpha,\ \beta\le ce\end{aligned} \tag{4}$$

where α and β are the nonnegative Lagrange multipliers. After optimizing the dual QPP (4), we obtain the weight vector

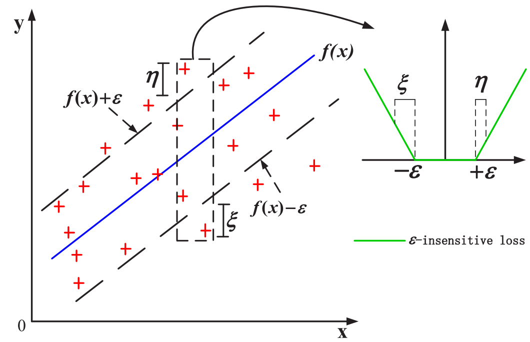

Fig. 1. The two-dimensional geometric interpretation and loss setting for ε-SVR.

$$w=A^T(\alpha-\beta) \tag{5}$$

Once we obtain the w, we can subsequently determine the bias term b by exploiting the Karush–Kuhn–Tucker (K.K.T) conditions [1]. Then the estimated regressor is constructed as follows:

$$f(x)=w^Tx+b \tag{6}$$

and the down- and up-bound of the regression model are constructed as follows:

$$f(x)-\varepsilon=w^Tx+b-\varepsilon \quad\text{and}\quad f(x)+\varepsilon=w^Tx+b+\varepsilon \tag{7}$$

2.2. ε-Twin support vector regression (ε-TSVR)

Different from ε-SVR, ε-TSVR finds a pair of ε-insensitive nonparallel proximal functions

$$f_1(x)=w_1^Tx+b_1 \quad\text{and}\quad f_2(x)=w_2^Tx+b_2 \tag{8}$$

Here, the empirical risks are measured by

$$R_{emp}^{\varepsilon_1}[f_1]=\sum_{i=1}^{m}\max\{0,(y_i-f_1(x_i))^2\}+c_1\sum_{i=1}^{m}\max\{0,-(y_i-f_1(x_i)+\varepsilon_1)\} \tag{9}$$

and

$$R_{emp}^{\varepsilon_2}[f_2]=\sum_{i=1}^{m}\max\{0,(f_2(x_i(-y_i)^2\}+c_2\sum_{i=1}^{m}\max\{0,-(f_2(x_i)-y_i+\varepsilon_2)\} \tag{10}$$

The functions $f_1(x)$ and $f_2(x)$ are obtained by solving the following pair of primal QPPs:

$$\begin{aligned}\min_{w_1,b_1,\xi}\quad & \frac{1}{2}c_3(w_1^Tw_1+b_1^2)+\frac{1}{2}\xi^{*T}\xi^*+c_1e^T\xi\\ \text{s.t.}\quad & Y-(Aw_1+eb_1)=\xi^*\\ & Y-(Aw_1+eb_1)\ge-\varepsilon_1e-\xi,\quad \xi\ge0\end{aligned} \tag{11}$$

and

$$\begin{aligned}\min_{w_2,b_2,\eta}\quad & \frac{1}{2}c_4(w_2^Tw_2+b_2^2)+\frac{1}{2}\eta^{*T}\eta^*+c_2e^T\eta\\ \text{s.t.}\quad & (Aw_2+eb_2)-Y=\eta^*\\ & (Aw_2+eb_2)-Y\ge-\varepsilon_2e-\eta,\quad \eta\ge0\end{aligned} \tag{12}$$

where c_1,c_2,c_3, and c_4 are positive parameters, ε_1 and ε_2 are nonnegative. For the optimization problem (11), the second term in the objective function is the sum of squared distances from the fitting function $f_1(x)=w_1^Tx+b_1$ to the training samples Y, where the least squares loss function are used. Therefore, minimizing it leads to the function $f_1(x)$ fit the regressor. The third term of the objective function, where the ε-insensitive loss function is used, minimizes the sum of error variables, thus attempting to over-fit the training samples. For the optimization problem (12), it has the