Neural Networks 74 (2016) 14–21

Contents lists available at ScienceDirect

Neural Networks

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

TWSVR: Regression via Twin Support Vector Machine

Reshma Khemchandani [a,*], Keshav Goyal [b], Suresh Chandra [b]

[a] Department of Computer Science, Faculty of Mathematics and Computer Science, South Asian University, Delhi, India

[b] Department of Mathematics, Indian Institute of Technology, Delhi, India

ARTICLE INFO

Article history:
Received 3 January 2015
Received in revised form 24 August 2015
Accepted 15 October 2015
Available online 3 November 2015

Keywords:
Machine Learning
Support Vector Machines
Support vector regression
Twin Support Vector Machines

ABSTRACT

Taking motivation from Twin Support Vector Machine (TWSVM) formulation, Peng (2010) attempted to propose Twin Support Vector Regression (TSVR) where the regressor is obtained via solving a pair of quadratic programming problems (QPPs). In this paper we argue that TSVR formulation is not in the true spirit of TWSVM. Further, taking motivation from Bi and Bennett (2003), we propose an alternative approach to find a formulation for Twin Support Vector Regression (TWSVR) which is in the true spirit of TWSVM. We show that our proposed TWSVR can be derived from TWSVM for an appropriately constructed classification problem. To check the efficacy of our proposed TWSVR we compare its performance with TSVR and classical Support Vector Regression(SVR) on various regression datasets.

1. Introduction

The last decade has witnessed the evolution of Support Vector Machines (SVMs) as a powerful paradigm for pattern classification and regression (Burges, 1998; Cherkassky & Mulier, 1998; Cortes & Vapnik, 1995). SVMs emerged from research in statistical learning theory on how to regulate the trade-off between structural complexity and empirical risk. The SVM classifier is also known as "maximum margin" classifier as it attempts to reduce the generalization error by maximizing the margin between two disjoint half planes (Burges, 1998; Cherkassky & Mulier, 1998; Cortes & Vapnik, 1995). The resultant optimization task involves the minimization of a convex quadratic function subject to linear inequality constraints. Support Vector Regression (SVR) is a technique for handling regression problems which is similar in principle to SVMs. The standard epsilon insensitive SVR model sets an epsilon tube around data points within which errors are discarded using an epsilon insensitive loss function.

Twin Support Vector Machine (TWSVM) (Jayadeva, Khemchandani, & Chandra, 2007) is a novel binary classification technique that determines two non parallel planes by solving two related SVM-type problems, each of which is smaller than in a conventional SVM. Recently, Peng (2010) proposed Twin Support Vector Regression (TSVR) which like TWSVM solves two quadratic programming problems (QPPs) to find two non-parallel regressor planes. Apparently the main reason for this terminology seems to be the belief that it is in the spirit of TWSVM. But we show in the sequel that this is not the case, because the formulation of Peng (2010) does not capture the essence of TWSVM.

In this paper, we provide a new framework of twin model to support vector regression problem, termed as TWSVR, which is in the true spirit to TWSVM. Unlike (Peng, 2010), our idea of twin support vector regression is truly inspired from TWSVM (Jayadeva et al., 2007), where the upper bound regressor(respectively lower bound regressor) problem deals with the proximity of points in upper tube(respectively lower tube) and at the same time at least ϵ distance from the points in the lower tube(respectively upper tube). Further, Bi and Bennett (2003) have developed an intuitive geometric framework for support vector regression(SVR) showing that SVR can be related to an appropriate SVM classification problem. Working on their line of work, we derive TWSVR formulation which is in the true spirit of TWSVM for classification problem. Though both of the aforementioned formulations solve two smaller sized QPPs instead of solving a large one as in classical SVR, only our proposed formulation can really be termed as Twin Support Vector Machine Based Regressor. We differentiate between the two notations TSVR and TWSVR. TSVR refers to Peng's formulation (Peng, 2010) whereas TWSVR refers to our proposed formulation. The terminology of TWSVR seems more natural because the basic twin classification formulation is termed as TWSVM.

The paper is organized as follows. Section 2 introduces the notations used in the rest of the paper and briefly discusses the formulations of Support Vector Machine and Twin Support Vector

* Corresponding author.
E-mail addresses: reshma.khemchandani@sau.ac.in (R. Khemchandani), gkeshav91@gmail.com (K. Goyal), chandras@maths.iitd.ac.in (S. Chandra).

http://dx.doi.org/10.1016/j.neunet.2015.10.007
0893-6080/

Machine. In Section 3, we first review Peng's model (Peng, 2010) and then point out reasons to conclude that formulation of Peng (2010) is not in the true spirit of TWSVM. Section 4 discusses the relevant result from Bi and Bennett (2003) where the equivalence between SVR and a related SVM is presented. Taking motivation from this equivalence of Bi and Bennett (2003), we extend the same to twin framework which leads to our TWSVR formulation in Section 5. In Section 6, we derive the dual formulation of TWSVR. Section 7 compares the results of TWSVR with TSVR and SVR on standard datasets. Section 8 provides concluding remarks.

2. Twin Support Vector Machine formulation

Let the samples to be trained be denoted by a set of m row vectors A_i, $i = 1, 2, \ldots, m$ in the n-dimensional real space $\mathbb{R}^n$, where the ith sample $A_i = (A_{i1}, A_{i2}, \ldots, A_{in})$. Also let $A = (A_1; A_2; \ldots; A_m)$ and $Y = (y_1; y_2; \ldots; y_m)$ denote the response vector of training samples. For classification problem $y_i \in \{1, -1\}$ and for regression problem $y_i \in \mathbb{R}$. Our aim is to find the parameters w and b, where $w \in \mathbb{R}^n$ and $b \in \mathbb{R}$, which in case of classification problem determines a classifier of the form $x^T w + b = 0$ and in case of regression problem determines a regressor function of the form $f(x) = x^T w + b$, where x is any point in n-dimension real space.

The Support Vector Machine(SVM) classifier is obtained by maximizing the margin between the bounding planes $x^T w + b = 1$ and $x^T w + b = -1$ along with minimizing the error associated with misclassified points with respect to bounding planes, and is equivalent to the following problem

$$
\text{(SVM)} \quad \underset{(w,\, b,\, q)}{Min} \quad C\, e^T q + \frac{1}{2} w^T w
$$
$$
s.t. \quad \begin{aligned} & A_i w + q_i \ge 1 - b \quad \text{for } y_i = 1, \\ & A_i w - q_i \le -1 - b \quad \text{for } y_i = -1, \\ & q_i \ge 0, \quad i = 1, 2, \ldots, m. \end{aligned} \tag{1}
$$

Here $q = [q_1, q_2, \ldots, q_m]^T$, where q_i denotes the margin error associated with the ith data sample, e is the vector of ones of appropriate dimensions and $C > 0$ denotes a scalar whose value determines the trade-off; a larger value of C emphasizes the classification error, while a smaller one places more importance on the classification margin.

In practice, rather than solving (SVM), we solve its dual problem to get the appropriate classifier. The formulations are extended on similar lines to handle nonlinear regression problems (Gunn, 1998).

Twin Support Vector Machine(TWSVM) (Jayadeva et al., 2007) tends to solve two QPPs of smaller size as opposed to a single large one in classical SVM making it almost four times faster than SVM in respect to training times.

The TWSVM classifier is obtained by solving the following pair of QPPs which yields two hyperplanes $x^T w^{(1)} + b^{(1)} = 0$ and $x^T w^{(2)} + b^{(2)} = 0$

$$
\text{(TWSVM1)} \quad \underset{(w^{(1)},\, b^{(1)},\, q^{(1)})}{Min} \quad \frac{1}{2}\|(Aw^{(1)} + e_1 b^{(1)})\|_2 + C_1 e_2^T q^{(1)}
$$
$$
s.t. \quad \begin{aligned} & -(Bw^{(1)} + e_2 b^{(1)}) + q^{(1)} \ge e_2, \\ & q^{(1)} \ge 0, \end{aligned} \tag{2}
$$

and,

$$
\text{(TWSVM2)} \quad \underset{(w^{(2)},\, b^{(2)},\, q^{(2)})}{Min} \quad \frac{1}{2}\|(Bw^{(2)} + e_2 b^{(2)})\|_2 + C_2 e_1^T q^{(2)}
$$
$$
s.t. \quad \begin{aligned} & (Aw^{(2)} + e_1 b^{(2)}) + q^{(2)} \ge e_1, \\ & q^{(2)} \ge 0, \end{aligned} \tag{3}
$$

where $C_1, C_2 > 0$ are parameters, $q^{(1)}, q^{(2)}$ denotes error variables, e_1 and e_2 are vectors of ones of appropriate dimensions, A and B are set of training examples corresponding to class 1 and -1 respectively.

The algorithm finds two hyperplanes, one for each class, and classifies points according to which hyperplane a given point is closest to. The first term in the objective function of (TWSVM1) or (TWSVM2) is the sum of squared distances from the hyperplane to points of one class. Therefore, minimizing it tends to keep the hyperplane close to points of one class (say class 1). The constraints require the hyperplane to be at a distance of at least 1 from points of the other class (say class -1); a set of error variables is used to measure the error wherever the hyperplane is closer than this minimum distance of 1. The second term of the objective function minimizes the sum of error variables, thus attempting to minimize mis-classification due to points belonging to class -1.

3. TSVR: Peng's model (Peng, 2010)

The TSVR formulations (Peng, 2010) solves the following two QPPs

$$
\text{(TSVR1)} \quad \underset{(w^{(1)},\, b^{(1)},\, \xi_1)}{Min} \quad \frac{1}{2}\|(Y - e\epsilon_1 - (Aw^{(1)} + eb^{(1)}))\|_2 + C_1 e^T \xi_1
$$
$$
s.t. \quad \begin{aligned} & Y - e\epsilon_1 - (Aw^{(1)} + eb^{(1)}) \ge -\xi_1, \\ & \xi_1 \ge 0, \end{aligned} \tag{4}
$$

and,

$$
\text{(TSVR2)} \quad \underset{(w^{(2)},\, b^{(2)},\, \xi_2)}{Min} \quad \frac{1}{2}\|(Y + e\epsilon_2 - (Aw^{(2)} + eb^{(2)}))\|_2 + C_2 e^T \xi_2
$$
$$
s.t. \quad \begin{aligned} & (Aw^{(2)} + eb^{(2)}) - (Y + e\epsilon_2) \ge -\xi_2, \\ & \xi_2 \ge 0, \end{aligned} \tag{5}
$$

where C_1, and $C_2 > 0$, ϵ_1, and $\epsilon_2 > 0$ are parameters, ξ_1, ξ_2 are slack vectors, e denotes vector of ones of appropriate dimension and $\|.\|_2$ denotes the L_2 norm.

Each of the above two QPP is smaller than the one obtained in the classical SVR formulation. Also (TSVR1) finds $f_1(x) = x^T w^{(1)} + b^{(1)}$ the down bound regressor and (TSVR2) finds the up bound regressor $f_2(x) = x^T w^{(2)} + b^{(2)}$. The final regressor is taken as the mean of up and down bound regressor.

We would now like to make certain remarks on the above formulations of TSVR. These remarks not only convince that formulation of Peng (2010) is not in the true spirit of TWSVM but also motivate us for our proposed formulation. In this context, we have the following lemma

Lemma 1. *For the given dataset, let $f(x)$ be the final regressor obtained from (TSVR1) and (TSVR2) when $\epsilon_1 = \epsilon_2 = 0$, and $g(x)$ be the final regressor obtained for any constant value of ϵ_1 and ϵ_2. Then*

$$g(x) = f(x) - (\epsilon_1 - \epsilon_2)/2.$$

Proof. Let $(w^{(1)}, b^{(1)})$ and $(w^{(2)}, b^{(2)})$ be the solutions to (TSVR1) and (TSVR2) respectively for constant ϵ_1 and ϵ_2 so that $g(x) = (x^T w^{(1)} + x^T w^{(2)})/2 + (b^{(1)} + b^{(2)})/2$. Now applying the transformation $b^{(1)}_{new} = b^{(1)} + \epsilon_1$ to (TSVR1) and $b^{(2)}_{new} = b^{(2)} - \epsilon_2$ to (TSVR2), we note that the resulting formulations have no ϵ term in it. The final regressor obtained from these transformed formulations will be $f(x)$. It follows from the transformation that $(w^{(1)}, b^{(1)} + \epsilon_1)$ and $(w^{(2)}, b^{(2)} - \epsilon_2)$ will be the solutions to transformed QPPs. Hence $f(x) = (x^T w^{(1)} + x^T w^{(2)})/2 + (b^{(1)} + b^{(2)})/2 + (\epsilon_1 - \epsilon_2)/2$, thus proving the result. □