Knowledge-Based Systems 160 (2018) 311–324

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Piecewise linear solution path for pinball twin support vector machine

Zhiji Yang[a,b], Xianli Pan[a], Yitian Xu[*,a]

[a] College of Science, China Agricultural University, Beijing 100083, China
[b] School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming 650221, China

ARTICLE INFO

Keywords:
Solution path
Piecewise linear
Support vector machine
Pinball loss

ABSTRACT

Twin support vector machine with pinball loss (PinTSVM) has been proposed recently, which enjoys noise insensitivity and has many admirable properties. However, users have to repeatedly train the model multiple times to tune parameters. To address this issue, in this paper we propose a new solution-path approach for the PinTSVM (Path-PinTSVM). We prove that both the primal and dual solutions are piecewise linear with the model parameters c and τ varying. The proposed algorithm could provide the optimal accuracy through all possible parameter values. The solution for the starting point of the path could be achieved analytically without solving optimization problem. Compared with the existing path algorithms for SVMs, our method is more flexible and has better prediction performance. As it deals with two classes separately, the analytic solution could be directly obtained no matter whether two classes are balanced or not. Besides, the computational cost is also less since only one class of the instances is considered at a time. Our approach also gives a guidance for exploiting path algorithms for other TSVMs. In numerical experiments, the validity of our proposed method is demonstrated on a synthetic dataset, 16 benchmark datasets and a real biological dataset.

1. Introduction

Since the well-known support vector machine (SVM) was proposed by Vapnik [1], many improved versions have been presented to enhance its prediction performance and computational efficiency [2]. Recently, a twin support vector machine (TSVM) was proposed by Jayadeva et al. [3] in the spirit of the proximal SVM [4]. The TSVM generates two nonparallel hyperplanes by solving a pair of smaller-sized problems such that each hyperplane is closer to one class and as far as possible from the other. The strategy of solving two smaller-sized quadratic programming problems (QPPs) rather than a single larger one makes the learning speed faster [5–13].

To further improve the performance of TSVM, the TSVM with pinball loss (PinTSVM) was proposed [14]. The PinTSVM also solves two smaller-sized QPPs. Its advantages are as follows: 1) by using pinball loss [15], it deals with the quantile distance and is less sensitive to noise points; 2) it is proved that the parameters could control bounds on proportions of boundary errors; 3) many admirable properties have been theoretically demonstrated, including between-class distance maximization and within-class scatter minimization. Although it has many advantages, the computational speed still needs to be improved. Besides, there are multiple parameters needed to be selected before training. One common way is the grid search approach. But it has two disadvantages: 1) the given value is not necessarily corresponding to the optimal classifier; 2) the repetitive training process of tuning parameters is time-consuming.

Recently, many approaches have been proposed to cope with this problem in SVMs [16–18]. As an attractive technique, the solution-path method has emerged, and it seeks to explore the entire solution path for all parameter values without having to re-train the model [19]. The goal of solution-path methods is to find the relationship between the solutions and parameters of the models. It has been proved that the solutions are piecewise-linear with the parameter c varying [20]. Moreover, the solution path has been applied to various machine learning tasks including support vector regression [21,22], the one-class SVM [23], the ranking SVM [24], ν-SVM [25,26], PinSVM [27] and other algorithms [16,28–33]. In addition, some researches on multi-dimensional solution paths have been studied [34–36]. These methods above are based on the model composed of one QPP, but so far there is no solution-path approach for solving TSVMs which are composed of a pair of QPPs. We notice that it could be more efficient to apply solution path into the twin model, since in TSVMs two classes of samples are considered separately, and for the initialization there is no need to consider whether the two classes are balanced or not.

In this paper, we prove that both the primal and dual solutions of the PinTSVM are piecewise linear as the parameter varies. Based on this

* Corresponding author.
E-mail address: xytshuxue@126.com (Y. Xu).

https://doi.org/10.1016/j.knosys.2018.07.022
Received 21 March 2018; Received in revised form 8 July 2018; Accepted 11 July 2018
Available online 20 July 2018
0950-7051/

property, we propose the solution-path algorithm. The entire solution paths for parameters c (Path-c-PinTSVM) and τ (Path-τ-PinTSVM) are given respectively. Whether two classes are balanced or not, the initial solutions could be represented in analytic forms. The overall computational complexity of the proposed method is much less than the traditional process of grid search.

Our main contributions are summarized as follows.

(i) The proposed algorithm could provide the optimal accuracy through all possible parameter values without repeatedly solving optimization problems.
(ii) By the strategy of dealing with two classes separately, the proposed algorithm is more flexible and runs faster than the traditional path algorithms of SVMs.
(iii) The solution paths for multiple parameters are presented, including the regularization parameter c and pinball loss parameter τ. In most of the exiting path algorithms, only one parameter is considered.
(iv) Our approach can be regarded as a guidance to exploit solution-path algorithms for other models of TSVMs. The property of piecewise linearity is suitable for other TSVMs.

The rest of this paper is organized as follows. Section 2 reviews the previous PinTSVM. In Section 3, the piecewise linearity of solutions is demonstrated. The solution paths for parameters c and τ are proposed in Sections 4 and 5, respectively. Section 6 performs numerical experiments on a synthetic dataset, 16 benchmark datasets and a real biological dataset to investigate the validity of the proposed algorithm. The last section contains the conclusions.

2. Review on pinball twin support vector machine

We consider a binary classification problem with a training set $T=\{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \cdots, (\boldsymbol{x}_l, y_l)\}$, where $\boldsymbol{x}_i \in R^n$ are inputs and $y_i \in \{1, -1\}$ are the corresponding outputs. The set contains l_1 positive instances and l_2 negative instances. We use column vectors $\boldsymbol{x}_i^+$ and $\boldsymbol{x}_j^-$ to denote positive and negative instances, respectively, where $i = 1, \cdots, l_1$ and $j = 1, \cdots, l_2$. Scalars are denoted by italic letters, and vectors by bold face letters. For vector $\boldsymbol{\lambda}$, let λ_i be the i^{th} component of $\boldsymbol{\lambda}$.

The PinTSVM, as an extension of TSVM, derives a pair of non-parallel planes by solving a pair of QPPs,

$$\begin{aligned} \min_{\boldsymbol{w}_+, b_+, \xi} \ & \frac{1}{2}\|\boldsymbol{w}_+\|^2 + \nu_1 \sum_{j=1}^{l_2} (\boldsymbol{w}_+^T \phi(\boldsymbol{x}_j^-) + b_+) + c_1 \sum_{i=1}^{l_1} \xi_i \\ \text{s.t.} \ & \boldsymbol{w}_+^T \phi(\boldsymbol{x}_i^+) + b_+ \geq -\xi_i, \\ & \boldsymbol{w}_+^T \phi(\boldsymbol{x}_i^+) + b_+ \leq \frac{1}{\tau}\xi_i, \ \ i = 1, 2\cdots, l_1, \end{aligned} \tag{1}$$

and

$$\begin{aligned} \min_{\boldsymbol{w}_-, b_-, \eta} \ & \frac{1}{2}\|\boldsymbol{w}_-\|^2 - \nu_2 \sum_{i=1}^{l_1} (\boldsymbol{w}_-^T \phi(\boldsymbol{x}_i^+) + b_-) + c_2 \sum_{j=1}^{l_2} \eta_j \\ \text{s.t.} \ & -(\boldsymbol{w}_-^T \phi(\boldsymbol{x}_j^-) + b_-) \geq -\eta_j, \\ & -(\boldsymbol{w}_-^T \phi(\boldsymbol{x}_j^-) + b_-) \leq \frac{1}{\tau}\eta_j, \ \ j = 1, 2, \cdots, l_2, \end{aligned} \tag{2}$$

where $\phi(\cdot)$ is a nonlinear mapping to a higher dimensional Hilbert space. By introducing Lagrangian function, we can derive their dual formulations as follows,

$$\begin{aligned} \max_{\lambda^+} \ & -\frac{1}{2}\sum_{i=1}^{l_1}\sum_{j=1}^{l_1} \lambda_i^+ K(\boldsymbol{x}_i^+, \boldsymbol{x}_j^+)\lambda_j^+ + \nu_1 \sum_{i=1}^{l_1}\sum_{j=1}^{l_2} K(\boldsymbol{x}_i^+, \boldsymbol{x}_j^-)\lambda_i^+ \\ \text{s.t.} \ & \sum_{i=1}^{l_1} \lambda_i^+ = \nu_1 l_2, \\ & -\tau_1 c_1 \leq \lambda_i^+ \leq c_1, \ i = 1, 2, \cdots, l_1, \end{aligned} \tag{3}$$

and

$$\begin{aligned} \max_{\lambda^-} \ & -\frac{1}{2}\sum_{i=1}^{l_2}\sum_{j=1}^{l_2} \lambda_i^- K(\boldsymbol{x}_i^-, \boldsymbol{x}_j^-)\lambda_j^- + \nu_2 \sum_{j=1}^{l_2}\sum_{i=1}^{l_1} K(\boldsymbol{x}_j^-, \boldsymbol{x}_i^+)\lambda_j^- \\ \text{s.t.} \ & \sum_{j=1}^{l_2} \lambda_j^- = \nu_2 l_1, \\ & -\tau_2 c_2 \leq \lambda_j^- \leq c_2, \ j = 1, 2, \cdots, l_2, \end{aligned} \tag{4}$$

where $\boldsymbol{\lambda}^+$ and $\boldsymbol{\lambda}^-$ are Lagrangian multipliers and kernel function is $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \langle \phi(\boldsymbol{x}_i), \phi(\boldsymbol{x}_j)\rangle$.

By solving QPPs (3) and (4), the Lagrangian vectors $\boldsymbol{\lambda}^+$ and $\boldsymbol{\lambda}^-$ could be achieved. Finally we can make the following decision function,

$$f(\boldsymbol{x}) = sgn\left(\frac{\boldsymbol{w}_+^T \phi(\boldsymbol{x}) + b_+}{\|\boldsymbol{w}_+\|} + \frac{\boldsymbol{w}_-^T \phi(\boldsymbol{x}) + b_-}{\|\boldsymbol{w}_-\|}\right). \tag{5}$$

3. Solution piecewise linearity in PinTSVM

In this section, we analyze the condition for achieving piecewise-linear solution. Then we prove that the solutions of the PinTSVM is piecewise linear, which is the foundation of our proposed algorithm.

3.1. The condition for piecewise-linear solution

Give an optimization problem as follows,

$$\min_{\boldsymbol{w}} g(\boldsymbol{w}) = J(\boldsymbol{w}) + cL(\boldsymbol{w}), \tag{6}$$

where $c > 0$ is the regularization parameter; J is a convex function and L is a convex nonnegative loss function.

For this problem, we provide the following lemma which gives a sufficient and necessary condition for the linearity of solution.

Lemma 1. *A sufficient and necessary condition for the solution path to be linear at c_0 when L, J are twice differentiable in a neighborhood of $\boldsymbol{w}(\boldsymbol{c_0})$ is that*

$$\frac{\partial \boldsymbol{w}(\boldsymbol{c})}{\partial \boldsymbol{c}} = -[\nabla^2 J(\boldsymbol{w}(\boldsymbol{c}) + c\nabla^2 L(\boldsymbol{w}(\boldsymbol{c})]^{-1} \nabla \boldsymbol{L}(\boldsymbol{w}(\boldsymbol{c}))$$

is a proportional (i.e., constant up to multiplication by a scalar) vector in $\mathbb{R}^l$ as a function of c in a neighborhood of c_0.

Proof. Since $\boldsymbol{w}$ is the optimal solution of the unconstrained optimization problem (6), the gradient function satisfies

$$\nabla \boldsymbol{g}(\boldsymbol{w}) = \nabla \boldsymbol{J}(\boldsymbol{w}) + c\nabla \boldsymbol{L}(\boldsymbol{w}) = 0. \tag{7}$$

For Eq. (7), regarding the optimal path $\boldsymbol{w}(\boldsymbol{c})$ as a function of c, we take a derivative with respect to c. Then we obtain

$$\nabla^2 J(\boldsymbol{w}(\boldsymbol{c}))\frac{\partial \boldsymbol{w}(\boldsymbol{c})}{\partial \boldsymbol{c}} + c\nabla^2 L(\boldsymbol{w}(\boldsymbol{c}))\frac{\partial \boldsymbol{w}(\boldsymbol{c})}{\partial \boldsymbol{c}} + \nabla \boldsymbol{L}(\boldsymbol{w}(\boldsymbol{c})) = 0.$$

If the matrix $\nabla^2 J(\boldsymbol{w}(\boldsymbol{c})) + c\nabla^2 L(\boldsymbol{w}(\boldsymbol{c}))$ is invertible, we can achieve

$$\frac{\partial \boldsymbol{w}(\boldsymbol{c})}{\partial \boldsymbol{c}} = -[\nabla^2 J(\boldsymbol{w}(\boldsymbol{c})) + c\nabla^2 L(\boldsymbol{w}(\boldsymbol{c}))]^{-1} \nabla \boldsymbol{L}(\boldsymbol{w}(\boldsymbol{c})).$$

Thus, the statement could be easily demonstrated. □

Lemma 1 implies sufficient conditions for piecewise linearity:

(i) J is piecewise quadratic as a function of $\boldsymbol{w}$ along the optimal path $\boldsymbol{w}(\boldsymbol{c})$;
(ii) L is piecewise linear as a function of $\boldsymbol{w}$ along this path.

We devote the next subsection to examining the PinTSVM which satisfies these conditions.