Journal of Computational Science 10 (2015) 270–277

Contents lists available at ScienceDirect

Journal of Computational Science

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

Twin support vector machine in linear programs

Dewei Li [a], Yingjie Tian [b,c,*]

[a] Information School, Renmin University of China, Beijing 100872, China
[b] Research Center on Fictitious Economy and Data Science, Chinese Academy of Sciences, Beijing 100190, China
[c] Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences, Beijing 100190, China

ARTICLE INFO

Article history:
Received 10 October 2014
Received in revised form 20 May 2015
Accepted 26 May 2015
Available online 30 May 2015

Keywords:
Twin support vector machine
Binary classification
Linear programs
Structural risk minimization

ABSTRACT

This paper propose a new algorithm, termed as LPTWSVM, for binary classification problem by seeking two nonparallel hyperplanes which is an improved method for TWSVM. We improve the recently proposed ITSVM and develop Generalized ITSVM. A linear function is chosen in the object function of Generalized ITSVM which leads to the primal problems of LPTWSVM. Comparing with TWSVM, a 1-norm regularization term is introduced to the objective function to implement structural risk minimization and the quadratic programming problems are changed to linear programming problems which can be solved fast and easily. Then we do not need to compute the large inverse matrices or use any optimization trick in solving our linear programs and the dual problems are unnecessary in the paper. We can introduce kernel function directly into nonlinear case which overcome the serious drawback of TWSVM. Also, we extend LPTWSVM to multi-class classification problem and get a new model MLPTWSVM. MLPTWSVM constructs M hyperplanes to make that the m-th hyperplane is far from the m-th class and close to the rest classes as much as possible which follow the idea of MBSVM. The numerical experiments verify that our new algorithms are very effective.

1. Introduction

Support vector machines (SVMs), as machine learning methods which were constructed on the VC-dimension theory and the principle of structural risk minimization, were proposed by Corinna Cortes and Vapnik in 1995 [1–3]. With the evolution of SVMs, they have shown much advantages in classification with small samples, nonlinear classification and high dimensional pattern recognition and also they can be applied in solving other machine learning problems [4–10]. The standard support vector classification attempts to minimize generalization error by maximizing the margin between two parallel hyperplanes, which results in dealing with an optimization task involving the minimization of a convex quadratic function. But some classifiers based on nonparallel hyperplanes were proposed recently. The generalized eigenvalue proximal support vector machine (GEPSVM) and twin support vector machine (TWSVM) are two typical classification methods and are also very popular. TWSVM seeks two nonparallel hyperplanes and make each hyperplane close to one class and far from the other as much as possible. However, the structural risk was not considered in TWSVM which may affect the computational efficiency and accuracy. Based on TWSVM, TBSVM and ITSVM was proposed in [11,12] which introduces a regularization term into the objective function and their experiments perform better than TWSVM. The main contribution of the twin bounded support vector machine (TBSVM) is that the principle of structural risk minimization is implemented by adding the regularization term in the primal problems. And the advantages of the improved twin support vector machine (ITSVM) are that it can apply kernel trick directly in the nonlinear case and it does not need to compute inverse matrices. Similar as SVM, a least squares version of TWSVM (LSTWSVM) has been presented which only require to solve two systems of linear equations [13]. LSTWSVM can get comparable generalization performance to TWSVM. Tian etc. proposed NPSVM to make TWSVM sparse by replace the quadratic loss function with $\epsilon-$insensitive loss function [14]. But the above mentioned improved algorithms all have their own drawbacks. TBSVM and LSTWSVM cannot introduce kernel function into linear case directly and LSTWSVM does not consider minimizing structural risk. ITSVM and NPSVM need to solve quadratic programs which need many complex tricks. TWSVM have been studied extensively [15–20].

Since multi-class classification problem is a natural extension of the binary classification problem, "One versus the rest" and "One versus one" [21,22] which construct a series of quadratic programs have been proposed and become two typical methods in solving multi-class classification problems. Recently, MBSVM [23] has been presented for multi-class classification inspired by the idea of TWSVM. For $M(M \geq 3)$ class classification, MBSVM seeks for M

* Corresponding author.
E-mail addresses: ruclidewei@126.com (D. Li), tyj@ucas.ac.cn (Y. Tian).

http://dx.doi.org/10.1016/j.jocs.2015.05.005
1877-7503/

hyperplanes such that each hyperplane is proximal to $M-1$ class and as far as possible from the rest one class. Because the decision criterion of MBSVM is the farthest distance of the test point to the hyperplanes, MBSVM have much lower computational complexity and can be solved faster than "One versus the rest" and "One versus one". However, MBSVM has two serious drawbacks which may affect its predict accuracy. It has to compute the inverse matrices though solving optimization problems which can lead to "Curse of Dimensionality" when the sample size is very large. Besides, it has not taken into account the confidence interval of structural risk so that its learning ability and generalization ability cannot meet requirements.

In this paper, we propose a novel approach to classification problem which involves two nonparallel hyperplanes in two linear optimization problems, termed as LPTWSVM, for binary classification. Since ITSVM is a successful method as an improved version of TWSVM, we develop Generalized ITSVM which follows the idea in [24]. Quadratic programs are time-consuming and need to be solved by so many optimization tricks. In comparison, linear programs are very popular because the formulations are nice which can be computed simpler [25]. So we consider using linear programming to obtain the best hyperplane parameters. LPTWSVM replaces the abstract function in the objective function of Generalized ITSVM with 1-norm terms and then we convert them to linear programs which can be solved easily and quickly and inherits the advantage of ITSVM. We can implement the principle of structural minimization and avoid computing inverse matrices. Also kernel function can be introduced to nonlinear case directly as the standard SVMs usually do. We also make an extension for LPTWSVM and solve multi-class classification at last.

The paper is organized as follows: Section 2 briefly introduces two algorithms, the original TWSVM and ITSVM; Section 3 proposes our new method LPTWSVM; we extended LPTWSVM to multi-class classification in Section 4; numerical experiments are implemented in Section 5 and concluding remarks are summarized in Section 6.

2. Background

In this section, we introduce the original TWSVM and its improved algorithm ITSVM.

2.1. TWSVM

Consider the binary classification problem with the training set

$$T=\{(x_1,+1),\cdots,(x_p,+1),(x_{p+1},-1),\cdots,(x_{p+q},-1)\}, \tag{2.1}$$

where $x_i \in R^n$, $i=1,\cdots,p+q$. Let $A=(x_1,\cdots,x_p)^{\mathrm{T}} \in R^{p\times n}$, $B=(x_{p+1},\cdots,x_{p+q})^{\mathrm{T}} \in R^{q\times n}$, and $l=p+q$.

TWSVM constructs the following primal problems

$$\min_{w_+,b_+,\xi_-} \frac{1}{2}(Aw_++e_+b_+)^{\mathrm{T}}(Aw_++e_+b_+)+c_1e_-^{\mathrm{T}}\xi_-, \tag{2.2}$$

$$s.t. -(Bw_++e_-b_+)+\xi_- \geq e_-, \xi_- \geq 0, \tag{2.3}$$

and

$$\min_{w_-,b_-,\xi_+} \frac{1}{2}(Bw_-+e_-b_-)^{\mathrm{T}}(Bw_-+e_-b_-)+c_2e_+^{\mathrm{T}}\xi_+, \tag{2.4}$$

$$s.t. (Aw_-+e_+b_-)+\xi_+ \geq e_+, \xi_+ \geq 0, \tag{2.5}$$

where c_i, i=1,2,3,4 are the penalty parameters, e_+ and e_- are vectors of ones, ξ_+ and ξ_- are slack vectors, $e_+, \xi_+ \in R^p$, e_-, $\xi_- \in R^q$. The decision function is denoted by

$$Class = arg\min_{k=-,+} |(w_k\cdot x)+b_k| \tag{2.6}$$

where $|\cdot|$ is the absolute value.

2.2. ITSVM

ITSVM is the abbreviation of improved twin support vector machine which changes the form of TWSVM and construct the following primal problems in linear case

$$\min_{w_+,b_+,\eta_+,\xi_-} \frac{1}{2}c_3(||w_+||^2+b_+^2)+\frac{1}{2}\eta_+^{\mathrm{T}}\eta_++c_1e_-^{\mathrm{T}}\xi_-, \tag{2.7}$$

$$s.t. Aw_++e_+b_+=\eta_+, \tag{2.8}$$

$$-(Bw_++e_-b_+)+\xi_- \geq e_-, \xi_- \geq 0, \tag{2.9}$$

and

$$\min_{w_-,b_-,\eta_-,\xi_+} \frac{1}{2}c_4(||w_-||^2+b_-^2)+\frac{1}{2}\eta_-^{\mathrm{T}}\eta_-+c_2e_+^{\mathrm{T}}\xi_+, \tag{2.10}$$

$$s.t. Bw_-+e_-b_-=\eta_-, \tag{2.11}$$

$$(Aw_-+e_+b_-)+\xi_+ \geq e_+, \xi_+ \geq 0, \tag{2.12}$$

where c_i, i=1,2,3,4 are the penalty parameters, e_+ and e_- are vectors of ones, ξ_+ and ξ_- are slack vectors, e_+, ξ_+, $\eta_+ \in R^p$, e_-, ξ_-, $\eta_- \in R^q$.

The following dual problems are considered to be solved

$$\max_{\lambda,\alpha} -\frac{1}{2}(\lambda^{\mathrm{T}} \quad \alpha^{\mathrm{T}})\hat{Q}(\lambda^{\mathrm{T}} \quad \alpha^{\mathrm{T}})^{\mathrm{T}}+c_3e_-^{\mathrm{T}}\alpha, \tag{2.13}$$

$$s.t. 0 \leq \alpha \leq c_1e_-, \tag{2.14}$$

and

$$\max_{\theta,\gamma} -\frac{1}{2}(\theta^{\mathrm{T}} \quad \gamma^{\mathrm{T}})\tilde{Q}(\theta^{\mathrm{T}} \quad \gamma^{\mathrm{T}})^{\mathrm{T}}+c_4e_+^{\mathrm{T}}\gamma, \tag{2.15}$$

$$s.t. 0 \leq \gamma \leq c_2e_+, \tag{2.16}$$

where

$$\hat{Q}=\begin{pmatrix} AA^{\mathrm{T}}+c_3I_+ & AB^{\mathrm{T}} \\ AB^{\mathrm{T}} & BB^{\mathrm{T}} \end{pmatrix}+E, \tag{2.17}$$

$$\tilde{Q}=\begin{pmatrix} BB^{\mathrm{T}}+c_4I_- & BA^{\mathrm{T}} \\ BA^{\mathrm{T}} & AA^{\mathrm{T}} \end{pmatrix}+E, \tag{2.18}$$

and I_+ is the $p\times p$ identity matrix, I_- is the $q\times q$ identity matrix, E is the $l\times l$ matrix with all entries equal to one. Thus a new point $x \in R^n$ is predicted to the class by (2.6) where

$$w_+=-\frac{1}{c_3}(A^{\mathrm{T}}\lambda+B^{\mathrm{T}}\alpha), b_+=-\frac{1}{c_3}(e_+^{\mathrm{T}}\lambda+e_-^{\mathrm{T}}\alpha), \tag{2.19}$$

$$w_-=-\frac{1}{c_4}(B^{\mathrm{T}}\theta-A^{\mathrm{T}}\gamma), b_-=-\frac{1}{c_4}(e_-^{\mathrm{T}}\theta-e_+^{\mathrm{T}}\gamma), \tag{2.20}$$

For the nonlinear case, after introducing the kernel function, the two corresponding problems in the Hilbert space H are

$$\min_{w_+,b_+,\eta_+,\xi_-} \frac{1}{2}c_3(||w_+||^2+b_+^2)+\frac{1}{2}\eta_+^{\mathrm{T}}\eta_++c_1e_-^{\mathrm{T}}\xi_-, \tag{2.21}$$

$$s.t. \Phi(A)w_++e_+b_+=\eta_+, \tag{2.22}$$

$$-(\Phi(B)w_++e_-b_+)+\xi_- \geq e_-, \xi_- \geq 0, \tag{2.23}$$

and

$$\min_{w_-,b_-,\eta_-,\xi_+} \frac{1}{2}c_4(||w_-||^2+b_-^2)+\frac{1}{2}\eta_-^{\mathrm{T}}\eta_-+c_2e_+^{\mathrm{T}}\xi_+, \tag{2.24}$$

$$s.t. \Phi(B)w_-+e_-b_-=\eta_-, \tag{2.25}$$

$$(\Phi(A)w_-+e_+b_-)+\xi_+ \geq e_+, \xi_+ \geq 0, \tag{2.26}$$

Their dual problems are constructed and can be solved directly.