Information Sciences 263 (2014) 22–35

Contents lists available at ScienceDirect

Information Sciences

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

Nonparallel hyperplane support vector machine for binary classification problems

Yuan-Hai Shao [a], Wei-Jie Chen [a], Nai-Yang Deng [b,*]

[a] Zhijiang College, Zhejiang University of Technology, Hangzhou 310024, PR China
[b] College of Science, China Agricultural University, Beijing 100083, PR China

ARTICLE INFO

Article history:
Received 3 October 2012
Received in revised form 28 September 2013
Accepted 7 November 2013
Available online 15 November 2013

Keywords:
Pattern recognition
Binary classification
Support vector machines
Proximal classifiers
Nonparallel hyperplanes

ABSTRACT

In this paper, we propose a nonparallel hyperplane support vector machine (NHSVM) for binary classification problems. Our proposed NHSVM is formulated by clustering the training points according to the similarity between classes. It constructs two nonparallel hyperplanes simultaneously by solving a single quadratic programming problem, and is consistent between its predicting and training processes - an essential difference that distinguishes it from other nonparallel SVMs. This proposed NHSVM has been analyzed theoretically and implemented experimentally. The results of experiments conducted using it on both artificial and publicly available benchmark datasets confirm its feasibility and efficacy, especially for "Cross Planes" datasets and datasets with heteroscedastic noise.

1. Introduction

Support vector machines (SVMs) are excellent kernel-based tools for binary data classification and regression [36,5,37,6]. They are based on the statistical learning theory and were introduced by Vapnik et al. [36]. SVMs have outperformed most other systems, such as artificial neural networks [42,41], in a variety of applications, including in a wide spectrum of research areas such as pattern recognition [2], text categorization [11,43], biomedicine [16], and financial regression [10,18]. The standard linear support vector classification (SVC) was the earliest SVM model. Its primal problem can be understood in the following way: Construct two parallel supporting hyperplanes such that, on the one hand, the band between the two parallel hyperplanes separates the two classes (the positive and negative data points) well, and on the other hand, the width of this band is maximized, leading to the introduction of the regularization term. Thus, the structural risk minimization principle is implemented. The final separating hyperlane is selected to be the "middle one" between the above two supporting hyperplanes.

In contrast to the classic SVC, in which two parallel supporting hyperplanes are constructed, let us now consider two nonparallel hyperplanes, e.g., the generalized eigenvalue proximal support vector machine (GEPSVM) [22] and the twin support vector machine (TWSVM) [12]. More precisely, both GEPSVM and TWSVM seek a pair of nonparallel proximal hyperplanes defined as

$$f_+(x) = w_+^\top x + b_+ = 0 \text{ and } f_-(x) = w_-^\top x + b_- = 0, \quad (1)$$

such that each hyperplane is close to one of two classes and keeps as far as possible away from the other class. The nonparallel hyperplanes are obtained by solving two generalized eigenvalue problems and two quadratic programming problems (QPPs) in GEPSVM and TWSVM respectively. Experimental results presented in [22,12] underscore their efficacy and their special abil-

* Corresponding author. Tel./fax: +86 010 62736265.
E-mail address: dengnaiyang@cau.edu.cn (N.-Y. Deng).

0020-0255/$ - see front matter
http://dx.doi.org/10.1016/j.ins.2013.11.003

ity in dealing with the "Cross Planes" datasets. Consequently, nonparallel SVMs with nonparallel hyperplanes, such as the twin bounded SVM [34,32], robust twin SVM [29], semi-supervised twin SVM [28], twin parametric-margin SVM (TPMSVM) [25,40], difference distance-based twin SVM [26,27] and twin support vector regression [35], have been studied extensively.

Logically, for a reasonable classification, the predicting and training processes should be consistent. However, this is not so in the nonparallel SVMs above [33]. In the predicting process of these SVMs, the class of a new test point x depends on the two distances between point x and the two hyperplanes defined in (1). However, in their training processes, neither the corresponding distances between the training points and the two hyperplanes defined in (1), nor their approximations, appear simultaneously. This disadvantage, compounded by the lack of consistency, becomes more serious in practical problems with heteroscedastic noise, in which the noise strongly depends on the location of the input x. This observation has motivated us to construct a nonparallel model with the desired consistency between the training and predicting processes.

In this paper, we propose a novel nonparallel hyperplane support vector machine (NHSVM) model. Our NHSVM belongs to the class of nonparallel proximal classifiers. However, in contrast to the above nonparallel SVM models in which the two nonparallel hyperplanes are constructed separately, inspired by clustering, our NHSVM constructs the two nonparallel hyperplanes simultaneously: the two classes are gathered in the two nonparallel hyperplanes respectively; and each hyperplane is closer to one class and far away from the other class to some extent. These two nonparallel hyperplanes are obtained by solving a single quadratic programming problem (QPP). Compared with the other nonparallel SVMs, our NHSVM is not only consistent between its predicting and training processes from a logical point of view, but also improves the classification accuracy and is suitable for many cases, such as "Cross Planes" datasets and datasets with heteroscedastic noise. Furthermore, in order to shorten the training time, an effective successive overrelaxation (SOR) technique is introduced to solve the QPP. The results of experiments conducted on several artificial and publicly available benchmark datasets confirm that our proposed NHSVM is feasible and effective, especially for "Cross Planes" datasets and datasets with heteroscedastic noise.

The remainder of this paper is organized as follows: Section 2 gives a brief overview of SVM and TWSVM. Section 3 presents our proposed NHSVM and related theoretical analysis. Section 4 discusses the results of experiments conducted. Finally, concluding remarks are given in Section 5.

2. Background

Consider a binary classification problem in the n-dimensional real space R^n. The set of training points is represented by $T = \{(x_i, y_i) | i = 1, 2, \ldots, m\}$, where $x_i \in R^n$ is the input and $y_i \in \{+1, -1\}$ is the corresponding output, for $i = 1, \ldots, m$. The m_1 inputs of Class $+1$ are organized as the matrix $X_+ \in R^{m_1 \times n}$, the m_2 inputs of Class -1 as the matrix $X_- \in R^{m_2 \times n}$, the m_1 outputs of Class $+1$ as a vector $Y_+ \in R^{m_1}$, and the m_2 outputs of Class -1 as a vector $Y_- \in R^{m_2}$.

2.1. Support vector classification

Standard linear support vector classification (SVC) [1,6] searches for a separating hyperplane defined as

$$f(x) = w^\top x + b = 0, \tag{2}$$

where $w \in R^n$ and $b \in R$. To measure the empirical risk, the soft margin loss function $\sum_{i=1}^{m} \max(0, 1 - y_i(w^\top x_i + b))$ is used. By introducing the regularization term $\frac{1}{2}\|w\|^2$ and the slack variable $\xi = (\xi_1, \ldots, \xi_m)$, the primal problem of SVC can be expressed as

$$\begin{aligned} &\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}\xi_i, \\ &\text{s.t. } y_i(w^\top x_i + b) \geqslant 1 - \xi_i, \quad \xi_i \geqslant 0,\ i = 1, \ldots, m, \end{aligned} \tag{3}$$

where $C > 0$ is a parameter. Note that the minimization of the regularization term $\frac{1}{2}\|w\|^2$ is equivalent to the maximization of the margin between two parallel supporting hyperplanes $w^\top x + b = 1$ and $w^\top x + b = -1$. Further, the structural risk minimization principle is implemented in this problem. When we obtain the optimal solution to (3), a new data point is classified as $+1$ or -1 according to whether the decision function, Class $i = \mathrm{sgn}(w^\top x + b)$, yields one or zero respectively.

In practice, rather than solving (3), we solve its dual problem to get the appropriate soft (or hard) margin classifier. The case of nonlinear kernels is handled along lines similar to that used for linear kernels [31].

2.2. Twin support vector machine

The linear twin support vector machine (TWSVM) [12,34] seeks a pair of nonparallel hyperplanes in R^n defined as in (1), such that each hyperplane is closer to the data points of one class and far from the data points of the other class to some extent. A new data point is assigned to Class +1 or -1 depending upon its proximity to the two nonparallel hyperplanes. In order to find the hyperplanes, the solutions to the following primal problems are required:

$$\begin{aligned} \min_{w_+, b_+, \xi_-} \quad & \tfrac{1}{2}\|X_+ w_+ + e_+ b_+\|^2 + c_1 e_-^\top \xi_-, \\ \text{s.t.} \quad & -(X_- w_+ + e_- b_+) + \xi_- \geqslant e_-, \quad \xi_- \geqslant 0, \end{aligned} \tag{4}$$