Expert Systems With Applications 54 (2016) 95–104

Contents lists available at ScienceDirect

Expert Systems With Applications

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

A second-order cone programming formulation for nonparallel hyperplane support vector machine

Miguel Carrasco[a], Julio López[b], Sebastián Maldonado[a,*]

[a] Facultad de Ingeniería y Ciencias Aplicadas, Universidad de los Andes, Mons. Álvaro del Portillo 12455, Las Condes, Santiago, Chile
[b] Facultad de Ingeniería, Universidad Diego Portales, Ejército 441, Santiago, Chile

ARTICLE INFO

Keywords:
Support vector classification
Nonparallel hyperplane SVM
Second-order cone programming

ABSTRACT

Expert systems often rely heavily on the performance of binary classification methods. The need for accurate predictions in artificial intelligence has led to a plethora of novel approaches that aim at correctly predicting new instances based on nonlinear classifiers. In this context, Support Vector Machine (SVM) formulations via two nonparallel hyperplanes have received increasing attention due to their superior performance. In this work, we propose a novel formulation for the method, Nonparallel Hyperplane SVM. Its main contribution is the use of robust optimization techniques in order to construct nonlinear models with superior performance and appealing geometrical properties. Experiments on benchmark datasets demonstrate the virtues in terms of predictive performance compared with various other SVM formulations. Managerial insights and the relevance for intelligent systems are discussed based on the experimental outcomes.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Support Vector Machine is one of the most popular tools used for prediction in intelligent systems. Its superior performance and flexibility are appealing virtues that lead to numerous extensions. SVM has proved to be very effective in various expert systems applications, such as medical diagnosis (Ríos & Erazo, 2016), churn prediction (Ali & Aritürk, 2014), and human resources analytics (Saradhi & Palshikar, 2011).

Recently, second-order cone programming (SOCP) has been used not only as an alternative optimization scheme for SVM (Debnath, Muramatsu, & Takahashi, 2005), but also to derive robust formulations that follow the SVM principle of maximum margin (Maldonado & López, 2014a; Nath & Bhattacharyya, 2007). The goal of such models is to construct one that correctly classifies most instances of each training pattern even for the worst distribution of the class-conditional densities with a given mean and covariance matrix. Such methods have proved to be very effective in terms of classification performance (Maldonado & López, 2014b).

On the other hand, there is a promising new stream of research that extends SVM to constructing two nonparallel hyperplanes in such a way that each one is close to one of the classes, and as far as possible from the other. The most popular approach is Twin SVM (Jayadeva, Khemchandani, & Chandra, 2007; Shao, Zhang, Wang, & Deng, 2011), while some other extensions, such as Nonparallel Hyperplane SVM (NH-SVM) (Shao, Chen, & Deng, 2014), have also been proposed in the literature, claiming successful results. Twin SVM splits the original problem into two smaller subproblems, and the two hyperplanes are constructed independently. In contrast, NH-SVM solves a single problem to obtain both classifiers simultaneously.

In this work, we propose a novel SVM-based method that extends the ideas of NH-SVM to second-order cones. The approach constructs two nonparallel classifiers, and represents each training pattern by an ellipsoid characterized by the mean and covariance of each class, instead of the reduced convex hulls used in NH-SVM. Originally developed for linear classifiers, the method is also adapted to construct nonlinear classification functions via the kernel trick. The use of ellipsoids for SVM modeling has been applied successfully in the context of expert systems (Czarnecki & Tabor, 2014).

This paper is organized as follows: in Section 2 we present the relevant SVM formulations for this work: Twin SVM, NH-SVM, and SOCP-SVM. The proposed method based on SOCP for Nonparallel Hyperplane SVM is described in Section 3. Experimental results using seven benchmark data sets are given in Section 4. Finally, Section 5 provides the main conclusions of this work, discussing managerial insights and addressing future developments in the context of expert and intelligent systems.

* Corresponding author. Tel: +56 2 26181874.
E-mail addresses: micarrasco@uandes.cl (M. Carrasco), julio.lopez@udp.cl (J. López), smaldonado@uandes.cl (S. Maldonado).

http://dx.doi.org/10.1016/j.eswa.2016.01.044
0957-4174/© 2016 Elsevier Ltd. All rights reserved.

2. Prior work in SVM classification

In this section, we discuss the relevant SVM formulations in this work: standard soft-margin SVM (Cortes & Vapnik, 1995), Twin SVM (Jayadeva et al., 2007; Shao et al., 2011), Nonparallel Hyperplane SVM (Shao et al., 2014), and SVM based on second-order cone programming (Nath & Bhattacharyya, 2007).

2.1. Soft-margin support vector machine

Given a set of training examples and their respective labels ($\mathbf{x}_i$, y_i), where $\mathbf{x}_i \in \Re^n$, $i = 1, \dots, m$ and $y_i \in \{-1, +1\}$, the soft-margin SVM formulation aims at finding a classification function of the form $f(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + b$ by solving the following quadratic programming problem (QPP):

$$\begin{aligned} \min_{\mathbf{w},b,\xi} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^m \xi_i \\ \text{s.t.} \quad & y_i \cdot (\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i, \qquad i = 1, \dots, m, \\ & \xi_i \geq 0, \qquad i = 1, \dots, m, \end{aligned} \tag{1}$$

where $\boldsymbol{\xi} \in \Re^m$ is a set of slack variables and $C > 0$ is a regularization parameter.

A non-linear classification function can be obtained via the Kernel Trick on the dual of Formulation (1) (Schölkopf & Smola, 2002). This kernel-based SVM formulation follows:

$$\begin{aligned} \max_{\boldsymbol{\alpha}} \quad & \sum_{i=1}^m \alpha_i - \frac{1}{2}\sum_{i,s=1}^m \alpha_i \alpha_s y_i y_s \mathcal{K}(\mathbf{x}_i, \mathbf{x}_s) \\ \text{s.t.} \quad & \sum_{i=1}^m \alpha_i y_i = 0, \\ & 0 \leq \alpha_i \leq C, \qquad i = 1, \dots, m, \end{aligned} \tag{2}$$

where $\boldsymbol{\alpha} \in \Re^m$ is the set of dual variables corresponding to the constraints in (1). In this work we use the *Gaussian kernel*, which usually lead to best empirical results (see e.g. Maldonado, Weber, and Basak (2011); Schölkopf and Smola. (2002)), and has the following form:

$$\mathcal{K}(\mathbf{x}_i, \mathbf{x}_s) = \exp\left(-\frac{||\mathbf{x}_i - \mathbf{x}_s||^2}{2\sigma^2}\right), \tag{3}$$

where σ is a parameter that controls the width of the kernel (Schölkopf and Smola. (2002)).

2.2. Twin support vector machine

The twin SVM performs classification by using two nonparallel hyperplanes obtained by solving two smaller-sized QPPs (Jayadeva et al., 2007). Let us denote the cardinality of the positive (negative) class by m_1 (m_2), and by $A \in \Re^{m_1 \times n}$ ($B \in \Re^{m_2 \times n}$) the data matrix related to the positive (negative) class. The linear Twin SVM formulation follows:

$$\begin{aligned} \min_{\mathbf{w}_1,b_1,\boldsymbol{\xi}_2} \quad & \frac{1}{2}\|A\mathbf{w}_1 + \mathbf{e}_1 b_1\|^2 + \frac{c_1}{2}(\|\mathbf{w}_1\|^2 + b_1^2) + c_3 \mathbf{e}_2^\top \boldsymbol{\xi}_2 \\ \text{s.t.} \quad & -(B\mathbf{w}_1 + \mathbf{e}_2 b_1) \geq \mathbf{e}_2 - \boldsymbol{\xi}_2, \\ & \boldsymbol{\xi}_2 \geq 0, \end{aligned} \tag{4}$$

and

$$\begin{aligned} \min_{\mathbf{w}_2,b_2,\boldsymbol{\xi}_1} \quad & \frac{1}{2}\| B\mathbf{w}_2 + \mathbf{e}_2 b_2\|^2 + \frac{c_2}{2}\left(\| \mathbf{w}_2\|^2 + b_2^2\right) + c_4 \mathbf{e}_1^\top \boldsymbol{\xi}_1 \\ \text{s.t.} \quad & (A\mathbf{w}_2 + \mathbf{e}_1 b_2) \geq \mathbf{e}_1 - \boldsymbol{\xi}_1, \\ & \boldsymbol{\xi}_1 \geq 0. \end{aligned} \tag{5}$$

Formulation (4)–(5) constructs two hyperplanes $\mathbf{w}_k^\top \mathbf{x} + b_k = 0$, $k = 1, 2$, such that each one is closer to instances of one of the two classes and is as far as possible from those of the other class. A new data point $\mathbf{x}$ is assigned to k^* according to its proximity to the hyperplanes based on the following rule:

$$k^* = \operatorname*{argmin}_{k=1,2}\left\{ d_k(\mathbf{x}) := \frac{|\mathbf{w}_k^\top \mathbf{x} + b_k|}{\|\mathbf{w}_k\|}\right\}, \tag{6}$$

where d_k corresponds to the perpendicular distance of the data sample $\mathbf{x}$ from hyperplane $\mathbf{w}_k^\top \mathbf{x} + b_k = 0$, $k = 1, 2$. The scalars c_1, c_2, c_3, and c_4 are positive parameters, and $\mathbf{e}_1$ and $\mathbf{e}_2$ are vectors of ones of appropriate dimensions. We refer to Formulation (4)–(5) as Twin-Bounded SVM (TB-SVM) (Shao et al., 2011), which extends the original Twin SVM (TW-SVM) formulation (Jayadeva et al., 2007). Both problems are equivalent if $c_1 = c_2 = \epsilon$, with $\epsilon > 0$ a fixed small parameter. The dual formulation of Twin-Bounded SVM can be found by Shao et al. (2011).

The linear Twin SVM can be extended to non-linear classification surfaces of the form $\mathcal{K}(\mathbf{x}, \mathbb{X})\mathbf{u}_k + b_k = 0$ $(k = 1, 2)$ via kernel functions by solving the following quadratic problems (kernel-based Twin SVM):

$$\begin{aligned} \min_{\mathbf{u}_1,b_1,\boldsymbol{\xi}_2} \quad & \frac{1}{2}\left\|\mathcal{K}(A^\top, \mathbb{X})\mathbf{u}_1 + \mathbf{e}_1 b_1\right\|^2 + \frac{c_1}{2}(\|\mathbf{u}_1\|^2 + b_1^2) + c_3 \mathbf{e}_2^\top \boldsymbol{\xi}_2 \\ \text{s.t.} \quad & -(\mathcal{K}(B^\top, \mathbb{X})\mathbf{u}_1 + \mathbf{e}_2 b_1) \geq \mathbf{e}_2 - \boldsymbol{\xi}_2, \\ & \boldsymbol{\xi}_2 \geq 0, \end{aligned} \tag{7}$$

and

$$\begin{aligned} \min_{\mathbf{u}_2,b_2,\boldsymbol{\xi}_1} \quad & \frac{1}{2}\left\|\mathcal{K}(B^\top, \mathbb{X})\mathbf{u}_2 + \mathbf{e}_2 b_2\right\|^2 + \frac{c_2}{2}(\|\mathbf{u}_2\|^2 + b_2^2) + c_4 \mathbf{e}_1^\top \boldsymbol{\xi}_1 \\ \text{s.t.} \quad & (\mathcal{K}(A^\top, \mathbb{X})\mathbf{u}_2 + \mathbf{e}_1 b_2) \geq \mathbf{e}_1 - \boldsymbol{\xi}_1, \\ & \boldsymbol{\xi}_1 \geq 0, \end{aligned} \tag{8}$$

where $\mathbb{X} = [A^\top B^\top] \in \Re^{n \times m}$ is the matrix that combines both training patterns sorted by class, and $\mathcal{K} : \Re^n \times \Re^n \to \Re$ is a kernel function (Schölkopf and Smola. (2002)).

2.3. Nonparallel hyperplane SVM (NH-SVM)

The NH-SVM approach constructs two nonparallel hyperplanes simultaneously by solving a single QPP. Similarly to Twin SVM, the linear NH-SVM formulation finds two hyperplanes in $\Re^n$ such that each classifier is close to one of the training patterns and is as far as possible from the other. The main difference compared to Twin SVM is that, since one single QPP is constructed, both hyperplanes are simultaneously optimized in the same formulation. The linear NH-SVM formulation follows:

$$\begin{aligned} \min_{\substack{\mathbf{w}_k,b_k,\xi_k \\ k=1,2}} \quad & \frac{1}{2}\left(\|A\mathbf{w}_1 + \mathbf{e}_1 b_1\|^2 + \|B\mathbf{w}_2 + \mathbf{e}_2 b_2\|^2\right) \\ & + \frac{c_1}{2}\left(\|\mathbf{w}_1\|^2 + b_1^2 + \|\mathbf{w}_2\|^2 + b_2^2\right) + \frac{c_2}{2}\left(\mathbf{e}_1^\top \xi_1 + \mathbf{e}_2^\top \xi_2\right) \\ \text{s.t.} \quad & A\mathbf{w}_1 + \mathbf{e}_1 b_1 - A\mathbf{w}_2 - \mathbf{e}_1 b_2 \geq \mathbf{e}_1 - \boldsymbol{\xi}_1, \\ & B\mathbf{w}_2 + \mathbf{e}_2 b_2 - B\mathbf{w}_1 - \mathbf{e}_2 b_1 \geq \mathbf{e}_2 - \boldsymbol{\xi}_2, \\ & \boldsymbol{\xi}_1 \geq 0, \boldsymbol{\xi}_2 \geq 0, \end{aligned} \tag{9}$$

where c_1, $c_2 > 0$ are regularization parameters (Shao et al., 2014). A point $\mathbf{x}$ in $\Re^n$ is assigned to class k^* by identifying the nearest hyperplane according to Eq. (6).

The computation of the Lagrangian and the Karush–Kuhn–Tucker (KKT) conditions leads to the following dual formulation for Problem (9):

$$\begin{aligned} \max_{\boldsymbol{\alpha}} \quad & \mathbf{e}^\top \boldsymbol{\alpha} - \frac{1}{2}\boldsymbol{\alpha}^\top \bar{A}^\top \left[(H^\top H + c_1 I)^{-1} + (G^\top G + c_1 I)^{-1}\right]\bar{A}\boldsymbol{\alpha}, \\ \text{s.t.} \quad & 0 \leq \boldsymbol{\alpha} \leq c_2 \mathbf{e}, \end{aligned}$$

Download English Version:

https://daneshyari.com/en/article/383258

Download Persian Version:

https://daneshyari.com/article/383258

Daneshyari.com