Expert Systems with Applications 41 (2014) 8211–8224

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Two ellipsoid Support Vector Machines

Wojciech Marian Czarnecki *, Jacek Tabor

Faculty of Mathematics and Computer Science, Jagiellonian University, prof. Stanislawa Lojasiewicza 6, 30-348 Krakow, Poland

ARTICLE INFO

Article history:
Available online 15 July 2014

Keywords:
Support Vector Machines
Mahalanobis distance
Classification
Data preprocessing

ABSTRACT

In classification problems classes usually have different geometrical structure and therefore it seems natural for each class to have its own margin type. Existing methods using this principle lead to the construction of the different (from SVM) optimization problems. Although they outperform the standard model, they also prevent the utilization of existing SVM libraries. We propose an approach, named 2eSVM, which allows use of such method within the classical SVM framework.

This enables to perform a detailed comparison with the standard SVM. It occurs that classes in the resulting feature space are geometrically easier to separate and the trained model has better generalization properties. Moreover, based on evaluation on standard datasets, 2eSVM brings considerable profit for the linear classification process in terms of training time and quality.

We also construct the 2eSVM kernelization and perform the evaluation on the 5-HT_{2A} ligand activity prediction problem (real, fingerprint based data from the cheminformatic domain) which shows increased classification quality, reduced training time as well as resulting model's complexity.

1. Introduction

Binary classification is a core problem in machine learning, relevant from both theoretical and practical perspectives. Over the past decade, Support Vector Machine (SVM) model (Cortes & Vapnik, 1995) gained a great interest due to its good mathematical formulation accompanied with a great number of empirical results. Several modifications were proposed, ranging from the modifications of norms used in the main SVM's equation (Zhang, 2004), through considering Bayesian treatment of the problem (Tipping, 2001) to generalization from original separating hyperplanes to hyperspheres and beyond (Le, Tran, Hoang, Ma, & Sharma, 2011). From the perspective of our paper of crucial importance are the generalizations which implement the idea that every class should have its own margin type: Twin Mahalanobis SVM (Peng & Xu, 2012) and Maxi-Min Margin Machine (M^4) (Huang, Yang, King, & Lyu, 2008).

Most of existing SVM modifications have shown their superiority over the classical method in some contexts and applications, however in practice, Vapnik's model (with later kernelization) is still the most commonly used. This is caused in particular by the fact that most SVM modifications require considerable amount of time and specialist knowledge to use, while Vapnik's model is implemented in most machine learning packages.

This is why in this paper we introduce a Two ellipsoid SVM (2eSVM) model which uses two distinct margins' types and allows easy implementation within the classical SVM framework. In fact we treat SVM as a Black Box, and perform only the pre- and post-processing of the data, see Fig. 1. Our approach allows not only the use of existing SVM libraries, but also gives the ability of careful comparison with the classical SVM modifications like Mahalanobis SVM.

The main idea behind Maxi-Min Margin Machine (M^4) (Huang et al., 2008) on which we based our ideas, is to seek for the hyperplane which simultaneously maximizes the size of different margins for classes X_- and X_+. The process of finding the maximal separating margin in the standard SVM algorithm can be seen as searching for the biggest radius r such that the sets

$$X_- + B(0,r) \quad \text{and} \quad X_+ + B(0,r)$$

are linearly separable, where $B(0,r)$ denotes the standard ball with radius r centered at zero. From the geometrical and practical point of view it is better to use two different hyperellipsoids $B_-(0,r)$ and $B_+(0,r)$ (balls in different metrics) fitted for each class separately. Consequently, one seeks for the maximal r such that the sets

$$X_- + B_-(0,r) \quad \text{and} \quad X_+ + B_+(0,r)$$

are linearly separable, see Fig. 2.

* Corresponding author. Tel.: +48 12 664 7556.
E-mail addresses: wojciech.czarnecki@uj.edu.pl (W.M. Czarnecki), jacek.tabor@uj.edu.pl (J. Tabor).

http://dx.doi.org/10.1016/j.eswa.2014.07.015
0957-4174/

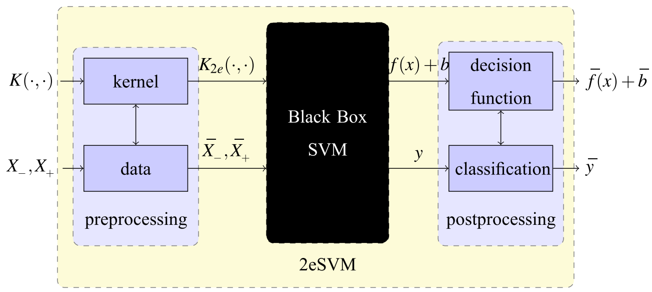


Fig. 1. Black Box scenario of 2`eSVM`.

As a result, the separating hyperplane is located nearer the "vertical" class. This is a better solution, as a big horizontal variance of the other class suggests that more points drawn randomly from the underlying distributions, which lay on the x axis "between" these two ellipsoids, are actually members of "horizontal" class.

This leads to the Second Order Cone Optimization Problem which cannot be solved by the standard SVM procedure (Huang et al., 2008). We prove however that we can implement a similar principle with the Black Box use of SVM by applying the following steps:

1. transform the data (or the kernel) using a matrix computed from the sum of classes' covariances,
2. train a classical SVM,
3. shift a decision boundary.

Evaluation on the typical datasets shows that the first operation allows the SVM to separate data easier and faster, while the third helps to obtain better classification results. In particular we obtain (see Section 6):

- a speed up of the learning process for the (both linear and kernelized) SVM of two to four times,
- reduction of mean number of support vectors of resulting model for the kernelized SVM by up to 10%,
- statistically significant better generalization results than the classical approach.

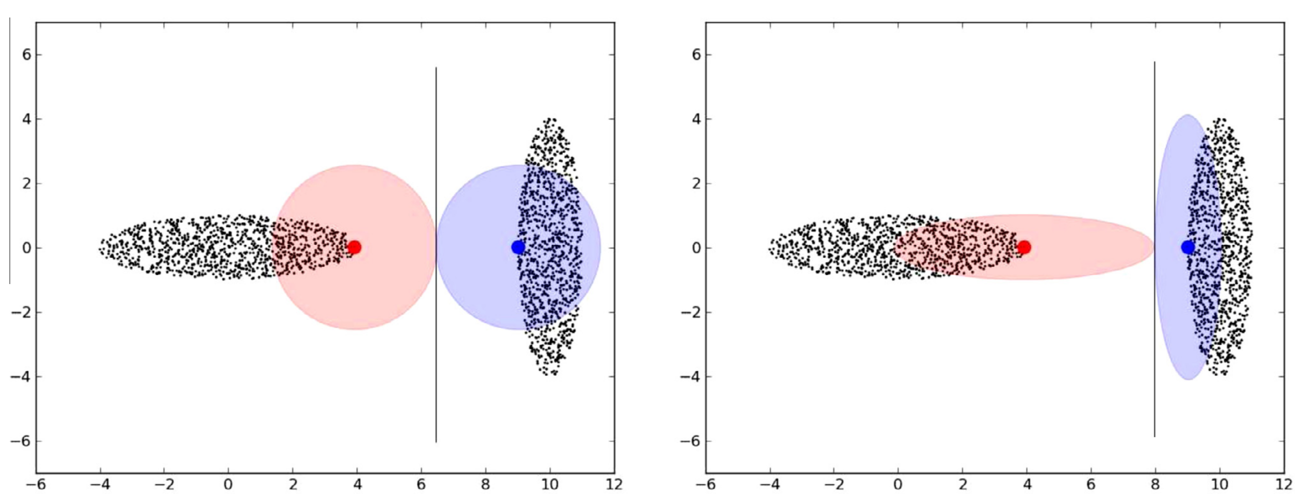
The most important differences between 2`eSVM` and M^4 are:

- 2`eSVM` is much simpler to implement, as it requires just few lines of algebraic operations,
- 2`eSVM` is much more robust, as it uses a SVM as an underlying optimization problem, which is a quadratic optimization with linear constraints, while M^4 requires second order cone optimization which is a much more complex optimization problem,
- M^4 requires custom optimization, while 2`eSVM` can be easily integrated with almost any existing SVM library,
- However, even though 2`eSVM` implements similar idea to M^4, due to its approximated nature, it achieves smaller accuracy gain.

The idea behind our method is related to the problem of finding the best metric as well as complexity reduction techniques. Metric learning concerns the problem of finding the best metric for given model and data as the independent optimization problem. Methods of this type have been used to build a hybrid model using both k-nearest neighbors and SVM concept – LM-KNN (Weinberger & Saul, 2009). There have also been presented multiple modifications of Support Vector Machines (Do, Kalousis, Wang, & Woznica, 2012), including incorporating the metric optimization in the core SVM optimization itself (Zhu, Gong, Zhao, & Zhang, 2012). The Ellipsoid SVM model (Momma, Hatano, & Nakayama, 2010) is a particular example of such approach, where one looks for the best fitted hyper-ellipsoid around the data to construct the correct metric. Those methods help model to better fit the underlying geometry of the data at the cost of additional computational requirements and in general – increased complexity of the problem.

On the other hand proximal SVM (Fung & Mangasarian, 2001) changes the basic formulation of the SVM to obtain a much simpler optimization problem. In this setting one searches for two parallel hyperplanes, around which points of particular classes are clustered, which are as far from each other as possible. Twin SVM (Jayadeva, Khemchandani, & Chandra, 2007) generalized this idea, so two hyperplanes can be non-parallel giving model better data geometrical fitting capabilities. Main strength of these methods lies in reduction of the complexity of the optimization problem by either weakening the parameter constraints (Fung & Mangasarian, 2001) or by solving multiple smaller problems (Jayadeva et al., 2007).

The proposed method differs from the above approaches, as it is based on the closed form of pre- and postprocessing methods of



(a) the same margin for both classes **(b)** each class has its own margin type

Fig. 2. Visualization of the idea on two elliptical-shaped classes (horizontal – positive samples and vertical – negative ones).