Applications of Support Vector Machines for Pattern Recognition: A Survey

Hyeran Byun[1] and Seong-Whan Lee[2]

[1]Department of Computer Science, Yonsei University
Shinchon-dong, Seodaemun-gu, Seoul 120-749, Korea
hrbyun@cs.yonsei.ac.kr
[2]Department of Computer Science and Engineering, Korea University
Anam-dong, Seongbuk-gu, Seoul 136-701, Korea
swlee@image.korea.ac.kr

Abstract. In this paper, we present a comprehensive survey on applications of Support Vector Machines (SVMs) for pattern recognition. Since SVMs show good generalization performance on many real-life data and the approach is properly motivated theoretically, it has been applied to wide range of applications. This paper describes a brief introduction of SVMs and summarizes its numerous applications.

1 Introduction

SVMs are a new type of pattern classifier based on a novel statistical learning technique that has been recently proposed by Vapnik and his co-workers [1-3]. Unlike traditional methods (e.g. Neural Networks), which minimize the empirical training error, SVMs aim at minimizing an upper bound of the generalization error through maximizing the margin between the separating hyperplane and the data [4]. Since SVMs are known to generalize well even in high dimensional spaces under small training sample conditions [5] and have shown to be superior to traditional empirical risk minimization principle employed by most of neural networks [6], SVMs have been successfully applied to a number of applications ranging from face detection, verification, and recognition [5-11,26,50-56,76-80,83], object detection and recognition [12-15,24,47,57], handwritten character and digit recognition [16-18,45], text detection and categorization [19,58-61], speech and speaker verification, recognition [20-23], information and image retrieval [33-36,87], prediction [37-41] and etc.[22,27-32,41,42,53,62,64,65,74].

In this paper, we aim to give a comprehensive survey on applications of SVMs for pattern recognition. This paper is organized as follows. We give a brief explanation on SVMs in Section 2 and a detailed review of SVMs-related techniques in Section 3. Section 4 describes the limitations of SVMs. We conclude this paper in Section 5.

S.-W. Lee and A. Verri (Eds.): SVM 2002, LNCS 2388, pp. 213-236, 2002.

2 Support Vector Machines

Classical learning approaches are designed to minimize error on the training dataset and it is called the Empirical Risk Minimization (ERM). Those learning methods follow the ERM principle and neural networks are the most common example of ERM. On the other hand, the SVMs are based on the Structural Risk Minimization (SRM) principle rooted in the statistical learning theory. It gives better generalization abilities (i.e. performances on unseen test data) and SRM is achieved through a minimization of the upper bound (i.e. sum of the training error rate and a term that depends on VC dimension) of the generalization error [1-3,43-45].

2.1 Linear Support Vector Machines for Linearly Separable Case

The basic idea of the SVMs is to construct a hyperplane as the decision plane, which separates the positive (+1) and negative (-1) classes with the largest margin, which is related to minimizing the VC dimension of SVM. In a binary classification problem where feature extraction is initially performed, let us label the training data $\mathbf{x}_i \in R^d$ with a label $y_i \in \{-1,+1\}$, for all the training data $i = 1,..., l$, where l is the number of data, and d is the dimension of the problem. When the two classes are linearly separable in R^d, we wish to find a separating hyperplane which gives the smallest generalization error among the infinite number of possible hyperplanes. Such an optimal hyperplane is the one with the maximum margin of separation between the two classes, where the margin is the sum of the distances from the hyperplane to the closest data points of each of the two classes. These closest data points are called Support Vectors (SVs). The solid line on Fig.1 represents the optimal separating hyperplane.

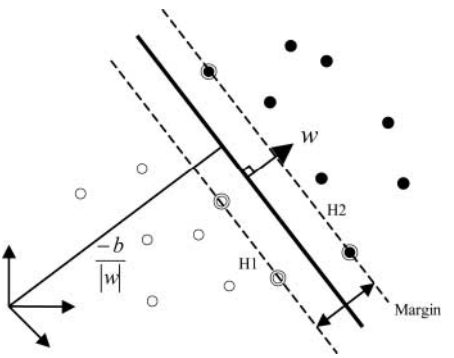

Fig. 1. Linear separating hyperplanes for the separable case. The support vectors are circled (taken from [44])