S

Support Vector Machine Models for Classification

Minghe Sun
College of Business, University of Texas at San Antonio, USA

INTRODUCTION

A support vector machine (SVM) is a quadratic programming (QP) model with training or learning algorithms. Developed by Vapnik (1995, 1998) and his coworkers, SVMs are used for classification, regression and function approximation. Because SVMs are QP models, this chapter discusses SVMs for classification from the mathematical programming (MP) perspective. SVMs are machine learning techniques because a SVM is usually trained, usually through numerical methods, to determine the values of the parameters in the classification function or discriminant functions.

As any other classification techniques, a SVM is used to construct a classification function or discriminant functions based on known values of the attributes or variables and known class memberships of the observations in a sample. The constructed classification function or discriminant functions are then used to evaluate the attribute values of any observation to obtain discriminant scores and to assign the observation into one of the classes. Many discriminant and classification techniques have been developed because no single technique always outperforms others under all situations (Johnson & Wichern, 1988). Statistical techniques, such as Fisher's linear discriminant function (Fisher, 1936), Smith's quadratic discriminant function (Smith, 1947) and logistic regression (Hand, 1981), have been standard tools for this purpose. More recently, other techniques, such as MP (Hand, 1981; Sun, 2010, 2011, 2014), neural networks (Stern, 1996) and classification trees (Breiman *et al*., 1984), have become alternative tools.

SVM (Vapnik, 1995, 1998) is a recent revolutionary development in classification analysis. Although they are modifications of the linear programming (LP) or mixed integer programming (MIP) models, they perform much better than their LP and MIP counterparts. Because they are QP models, they are also much more difficult to solve than their LP counterparts. Three concepts, classification margin maximization, dual formulation and kernels, are crucial in SVMs (Bredensteiner & Bennett, 1999). By minimizing the sum of classification errors and maximizing the classification margin at the same time in a QP model, SVMs construct discriminant functions with good generalization capabilities. Usually the dual formulation of the QP models is solved because the dual is usually easier to solve than the primal. By using inner product kernels in the dual formulations, SVMs can be built and nonlinear discriminant functions can be constructed in high dimensional feature spaces without carrying out the mappings from the input space to the high dimensional feature spaces. The size of the dual formulation is independent of the dimension of the input space and independent of the inner product kernels used.

However, most of the SVM research is for two-class classification although efforts have been made to extend the techniques to multi-class problems. Some of the approaches proposed in the literature for multi-class classification include the one-against-one approach (e.g., Friedman, 1996; Kreβel, 1999; Mayoraz & Alpaydin, 1999; Angulo *et al*., 2003), the one-against-all approach (e.g., Corte & Vapnik, 1995; Vapnik, 1995, 1998) and the one model formulation (e.g., Vapnik, 1995, 1998; Corte & Vapnik, 1995; Bredensteiner &

DOI: 10.4018/978-1-4666-5202-6.ch215

Copyright © 2014, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

Bennett, 1999; Weston & Watkins, 1999; Guermeur *et al*., 2000; Crammer & Singer, 2001; Guermeur, 2002; Lee *et al*., 2004; Sun, 2013). Examples of popular training software packages for SVMs include LIBSVM (Chang & Lin, 2011) and LIBLINEAR (Fan *et al*., 2008).

Two SVM models for discriminant and classification analysis are discussed in this chapter, one for two-class classification and the other for multi-class classification. Both the primal and dual formulations are discussed and an example is presented for each model. However, because SVMs are usually built in high dimensional feature spaces and usually the dual rather than the primal is solved, this chapter focuses on the dual formulations of SVM models in feature spaces using kernels.

CLASSIFICATION USING KERNELS

Assume a dataset with m observations in p classes is available for analysis. The index set of the classes is represented by K. The index set of the observations in the whole dataset is represented by I while that in class k is represented by I_k, for $1 \le k \le p$, such that $I = \bigcup_{k=1}^{p} I_k$. Similarly, the number of observations in class k is represented by m_k such that $m = \sum_{k=1}^{p} m_k$. Each observation is measured by an input vector of n variables. The variables represent the measurements on the characteristics or attributes of the observations. The input vector for a specific observation $i \in I$ in the dataset is represented by $\mathbf{x}_i \in \Re^n$ and that of a generic observation by $\mathbf{x} \in \Re^n$.

Oftentimes, nonlinear classification or discriminant functions perform better than linear ones. The observations in the classes are more accurately separated by a nonlinear classification function or by nonlinear discriminant functions than by linear ones. Therefore, nonlinear terms of the input variables, such as monomials and logarithms, are used in the classification or discriminant functions. The input vectors $\mathbf{x} \in \Re^n$ are mapped from the input space to feature vectors in a high dimensional feature space through the nonlinear mapping $\boldsymbol{\varphi}(\mathbf{x}) \in \Re^{n'}$, with $n' \gg n$. Hence, the classification function or the discriminant functions are constructed in the high dimensional feature space $\Re^{n'}$.

However, $\boldsymbol{\varphi}(\mathbf{x})$ appears in the dual formulation of the SVM models only in the form of inner products. Through the use of inner product kernels $K(\mathbf{x}_i, \mathbf{x}) = \boldsymbol{\varphi}'(\mathbf{x}_i)\boldsymbol{\varphi}(\mathbf{x})$, the nonlinear mappings are not necessarily carried out or the function forms of $\boldsymbol{\varphi}(\mathbf{x})$ are not necessarily known. Different inner product kernels have been used in SVMs (Vapnik, 1995, 1998), such as the polynomial kernel and the radial basis function (RBF), also called the Gaussian, kernel. The polynomial kernel has the form of

$$K(\mathbf{x}_i, \mathbf{x}) = (\mathbf{x}_i^{\mathrm{T}}\mathbf{x} + 1)^q, \tag{1}$$

where $q \ge 1$ is an integer. The RBF kernel has the form of

$$K(\mathbf{x}_i, \mathbf{x}) = \exp(-\gamma \mid \mathbf{x}_i - \mathbf{x} \mid^2), \tag{2}$$

where γ is a user provided parameter that may also be determined in the training process of the SVM. The polynomial and the RBF kernels will be used as examples in the following discussions. If a SVM is directly built in the input space, the mapping $\boldsymbol{\varphi}(\mathbf{x}) = \mathbf{x}$ is used.

When $p = 2$, the classification function in the feature space is of the form

$$f(\mathbf{x}) = b_0 + \mathbf{b}'\varphi(\mathbf{x}), \tag{3}$$

where $b_0 \in \Re$ and $\mathbf{b} \in \Re^{n'}$ are the estimated parameters. The value of $f(\mathbf{x})$ evaluated at a specific input vector $\mathbf{x}$ is called a classification score. A hyperplane represented by $f(\mathbf{x}) = 0$ is supposed to separate the $p = 2$ classes. The function $f(\mathbf{x})$ in (3) is a classification function

13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/support-vector-machine-models-for-classification/107423

Related Content

Exploration of Soft Computing Approaches in Itemset Mining
Jyothi Pillaiand O. P. Vyas (2016). *Business Intelligence: Concepts, Methodologies, Tools, and Applications (pp. 1830-1856).*
www.irma-international.org/chapter/exploration-of-soft-computing-approaches-in-itemset-mining/142705

Neural Data Mining System for Trust-Based Evaluation in Smart Organizations
T. T. Wong (2006). *Integration of ICT in Smart Organizations (pp. 159-186).*
www.irma-international.org/chapter/neural-data-mining-system-trust/24065

Towards Automation of Business Intelligence Services Using Hybrid Intelligent System Approach
Rajendra M. Sonar (2013). *International Journal of Business Intelligence Research (pp. 61-92).*
www.irma-international.org/article/towards-automation-of-business-intelligence-services-using-hybrid-intelligent-system-approach/104739

Kinked Experience Curve
Yu Sang Changand Jinsoo Lee (2014). *Encyclopedia of Business Analytics and Optimization (pp. 1358-1366).*
www.irma-international.org/chapter/kinked-experience-curve/107331

Organizational Sustainability: Aspects of Agility
Hakikur Rahman (2014). *International Journal of Business Intelligence Research (pp. 17-38).*
www.irma-international.org/article/organizational-sustainability/120050