Statistics and Probability Letters 132 (2018) 107–115

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Varying Coefficient Support Vector Machines

Xiaoling Lu [a,*], Fengchi Dong [a], Xiexin Liu [b], Xiangyu Chang [c]

[a] *Center for Applied Statistics, Data Mining Center, School of Statistics, Renmin University of China, Beijing, 100872, China*
[b] *Department of Management Sciences, Tippie College of Business, the University of Iowa, IA, 52242, USA*
[c] *Center of Data Science and Information Quality, School of Management, Xi'an Jiaotong University, Xi'an, 710049, China*

ARTICLE INFO

Article history:
Received 27 January 2016
Received in revised form 19 May 2017
Accepted 14 September 2017
Available online 7 October 2017

Keywords:
Support Vector Machines
Varying Coefficient model
Document classification

ABSTRACT

This paper proposes a Varying Coefficient Support Vector Machine (VCSVM). VCSMO, a variation of the classic algorithm SMO for standard SVMs, is also proposed to solve for VCSVM. Numerical examples validate the accuracy and efficiency of the proposed model.

1. Introduction

Classification remains to be an intriguing topic in the field of statistical machine learning. There are various available classification techniques, including logistic regression (Cox, 1958), decision trees (Breiman et al., 1984), support vector machines (Cortes and Vapnik, 1995), etc. Among all the classification techniques, max-margin classifiers like SVMs have been widely recognized (Ladicky and Torr, 2011). For a linear-separable sample, SVM finds its linear decision surface to maximize its margin; and for a non-linear separable problem, SVM maps the input matrix to a very high-dimensional feature space and constructs a linear decision surface in that space. For a non-separable problem, SVM penalizes training examples that are either misclassified or not far enough from the decision surface, and the optimization problem could be treated as a Quadratic Programming (QP) problem. Various algorithms have been proposed to solve SVMs (Shawe-Taylor and Sun, 2011). Sequential Minimal Optimization (SMO) (Platt, 1998) is a classic one. LIBSVM, the most popular library for SVM by far, also implements a SMO-type algorithm (Wu and Kumar, 2009; Chang and Lin, 2011).

In classification problems, it is common to observe that some coefficients of covariates vary over a certain index variable. For example, when we are trying to discriminate HIV carriers based on his/her symptoms, we need to consider the fact that the relationship between symptoms and results may change for people with different ages. In document classification problems, the importance of some keywords could also change at different time points. Nevertheless, though this is a common phenomenon in real world, it has not been thoroughly studied.

Under regression settings, varying coefficient models (Cleveland et al., 1991; Hastie and Tibshirani, 1993) could solve the above mentioned problem by setting the coefficient as a smooth function of a certain index variable. Cai et al. (2000) propose the varying coefficient logistic regression (VCLR) model which can be used for classification problem. In this article, we novelly employ the idea behind varying coefficient models, combine it with SVMs, and name it Varying Coefficient Support Vector Machine (VCSVM for short). The proposed VCSVM could, therefore, solve the above mentioned classification problems by replacing the constant coefficient in the classifier with a coefficient function. We propose a new algorithm

* Corresponding author.
E-mail address: xiaolinglu@ruc.edu.cn (X. Lu).

https://doi.org/10.1016/j.spl.2017.09.006
0167-7152/

Varying Coefficient Sequential Minimal Optimization (VCSMO), based on the classic algorithm SMO for training standard SVMs, in order to solve for VCSVM. In simulation studies, we compare the proposed method with VCLR and standard SVMs with different kernels in both low-dimension and high-dimension situations. The results show VCSVM and VCLR not only inherit the high interpretability of varying coefficient models, but also are able to achieve higher classification accuracy than SVMs. Furthermore, when the data contains more noises, VCSVM tends to perform better than VCLR. A real data analysis in classification of news further proves the strong applicability of the proposed model. Especially, VCSVM is preferred to VCLR because it has the property of sample sparsity, which means that it can select some typical documents to better summarize the data.

The rest of this paper is organized as follows. We first review varying coefficient models and SVMs, and propose VCSVM in Section 2. The detailed algorithm VCSMO to solve VCSVM is then presented in Section 3. Several simulation studies and real data analysis are completed to verify the proposed model in Sections 4 and 5. Section 6 concludes this paper.

2. Methodology

In this section, we first review varying coefficient models and their estimation methods, then we introduce the proposed VCSVM model.

2.1. Varying coefficient models

Let (x_i, y_i, z_i) be the ith observation for a training set $(x_1, y_1, z_1), \dots, (x_n, y_n, z_n)$, where the input $x_i \in \mathcal{R}^p$, $y_i \in \mathcal{R}$ is the corresponding continuous dependent variable, and z is the so-called index variable. A single index varying coefficient model then assumes

$$y_i = \beta_0(z_i) + x_{i1}\beta_1(z_i) + \cdots + x_{ip}\beta_p(z_i) + \epsilon_i \tag{2.1}$$

with ϵ_i being the random error satisfying $E(\epsilon_i|x_i, z_i) = 0$. $\boldsymbol{\beta}(z) = \{\beta_0(z), \beta_1(z), \dots, \beta_p(z)\}^\top \in \mathcal{R}^{p+1}$ is the coefficient vector with respect to z, which is usually assumed as smooth function. Varying coefficient models are recognized for their flexibility in that the model coefficients could vary with the index variable z. In real problems, z could be set as time, age, and others.

According to Fan and Zhang (2008), there are three approaches to estimate $\beta(z)$, including kernel-local polynomial smoothing, polynomial spline and smoothing spline. Among all three methods, here we choose kernel methods for its practicability. A locally kernel estimator $\hat{\beta}(z)$ could be obtained via

$$\min_{\beta,\beta_0} \sum_{i=1}^{n} \left(y_i - \left(\beta_0 + x_i^\top \beta\right)\right)^2 K_h(z - z_i), \tag{2.2}$$

where $K_h(\cdot) = K(\cdot/h)/h$ is the kernel function and h is its corresponding bandwidth. Common kernel functions include Epanechnikov kernel, Gaussian kernel, Uniform kernel and many others.

Cai et al. (2000) extend the above ideas to generalized linear model situation through link functions and propose the varying coefficient logistic regression model which can be used for classification problem.

2.2. Varying coefficient support vector machine

For a given training set $(x_1, y_1), \dots, (x_n, y_n)$, where the input $x_i \in \mathcal{R}^p$, and the corresponding response variable $y_i \in \{1, -1\}$, for a binary classification problem, a standard SVM minimizes

$$\begin{aligned} \min_{\beta,\beta_0} \quad & C\sum_{i=1}^{n} \xi_i + \frac{1}{2}\|\beta\|^2, \\ s.t. \quad & \xi_i \ge 0, \\ & y_i f(x_i) \ge 1 - \xi_i, \end{aligned} \tag{2.3}$$

where C is the cost parameter, and $f(x_i) = x_i^\top \beta + \beta_0$ is the classifier with constant coefficients β_0 and $\beta = (\beta_1, \dots, \beta_p)$.

When constant coefficients are not enough, we follow the idea behind varying coefficient models and propose Varying Coefficient Support Vector Machine (VCSVM). That is, for a training set $(x_1, y_1, z_1), \dots, (x_n, y_n, z_n)$, where the input $x_i \in \mathcal{R}^p$, the corresponding binary response variable is $y_i \in \{1, -1\}$, and z is the index variable with index $t = 1, \dots, n$, VCSVM solves the following Eq. (2.4) at each given index t.

$$\begin{aligned} \min_{\beta_0(z_t),\beta(z_t),\xi_{it}} \quad & C\sum_{i=1}^{n} \xi_{it} + \frac{1}{2}\|\beta(z_t)\|^2, \\ s.t. \quad & \xi_{it} \ge 0, \\ & \xi_{it} \ge [1 - y_i(\beta_0(z_t) + \beta(z_t)^T x_i)]_+ K_h(z_t - z_i). \end{aligned} \tag{2.4}$$