Neurocomputing 161 (2015) 254–259

Contents lists available at ScienceDirect

Neurocomputing

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

Varying coefficient modeling via least squares support vector regression

Jooyong Shim [a], Changha Hwang [b,*]

[a] Institute of Statistical Information, Department of Data Science, Inje University, Kyungnam 621-749, South Korea
[b] Department of Applied Statistics, Dankook University, Gyeonggido 448-160, South Korea

A R T I C L E I N F O

Article history:
Received 17 July 2014
Received in revised form
13 February 2015
Accepted 14 February 2015
Communicated by W.S. Hong
Available online 24 February 2015

Keywords:
Confidence interval
Generalized cross validation
Least squares support vector regression
Model selection
Varying coefficient model

A B S T R A C T

The varying coefficient regression model has received a great deal of attention as an important tool for modeling the dynamic changes of regression coefficients in the social and natural sciences. Lots of efforts have been devoted to develop effective estimation methods for such regression model. In this paper we propose a method for fitting the varying coefficient regression model using the least squares support vector regression technique, which analyzes the dynamic relation between a response and a group of covariates. We also consider a generalized cross validation method for choosing the hyperparameters which affect the performance of the proposed method. We provide a method for estimating the confidence intervals of coefficient functions. The proposed method is evaluated through simulation and real example studies.

1. Introduction

The varying coefficient model introduced by [4] is flexible and powerful for modeling the dynamic changes of regression coefficients. The varying coefficient model is a useful extension of the classical linear model. In this model, the regression coefficients are not set to be constants but are allowed to change smoothly with the value of other covariates. The varying coefficient model inherits simplicity and easy interpretation of the classical linear models. This model is gaining its popularity in statistics literature in recent years. The introductions, various applications and current research areas of the varying coefficient model can be found in [4,5,3,11]. Many areas of applied statistics have become aware of the problem of estimating varying coefficients and analyzing them appropriately. A great deal of attention has been focused on the problem of estimating the varying coefficients. Most of this attention has been paid to using the kernel smoothing technique. Fan and Zhang [3] give an excellent review of the varying coefficient models and discuss three approaches in estimating the coefficient function: kernel smoothing, polynomial splines and smoothing splines. Recently, some more flexible varying coefficient models have been developed and discussed. See, for example, [19,8,6,7,18].

There are some possibilities in building the varying coefficient model. One is to let all regression coefficients be functions of a single covariate or a single vector of covariates. Another is to let each regression coefficient be function of different covariates. There are various extensions of models. See [11] for details. Most literature has centered around the case that all regression coefficients are functions of a single vector of covariates. In this paper we consider this model. The desire to estimate the coefficients nonparametrically leads to the subject of this paper. In this paper we propose a method for fitting the varying coefficient model by utilizing least squares support vector regression (LS-SVR) technique, which can be applied effectively to high-dimensional case. From now on this method will be simply called VC-LS-SVR. This is the first paper utilizes the idea of LS-SVR or support vector machine (SVM) for the varying coefficient model. The SVM, first developed by [16] and his group at AT& T Bell Laboratories, has been successfully applied to a number of real world problems related to classification and regression problems. Least squares SVM (LS-SVM) is least squares version of SVM and was initially introduced by [14]. LS-SVM has been proved to be a very appealing and promising method [14,15]. There are some strong points of LS-SVM. Here we will consider three of them. One is that LS-SVM uses the linear equation which is simple to solve and good for computational time saving. Another is that LS-SVM makes the model selection easier by using the generalized cross validation (GCV) function. The other is that LS-SVM enables to construct an approximate pointwise confidence interval for the true regression function.

* Corresponding author.
E-mail addresses: ds1631@hanmail.net (J. Shim), chwang@dankook.ac.kr (C. Hwang).

http://dx.doi.org/10.1016/j.neucom.2015.02.036
0925-2312/

The rest of this paper is organized as follows. Section 2 proposes VC-LS-SVR method and also presents a GCV technique in order to choose the hyperparameters in the proposed method. Section 3 describes a method for estimating the confidence intervals of coefficient functions. Sections 4 and 5 present numerical studies and conclusion, respectively.

2. VC-LS-SVR and model selection

In this section we illustrate the VC-LS-SVR method and its model selection. The underlying idea of the VC-LS-SVR method is that the true mean specification is approximated by a combination of linear LS-SVR and nonlinear feature mapping function of the univariate or multivariate smoothing variable.

2.1. VC-LS-SVR

Given the training data set $\mathcal{D}=\{(\boldsymbol{u}_i,\boldsymbol{x}_i,y_i)\}_{i=1}^n$ with each smoothing vector $\boldsymbol{u}_i\in\mathbb{R}^{d_u}$, covariate vector $\boldsymbol{x}_i\in\mathbb{R}^{d_x}$ and the response $y_i\in\mathbb{R}$, we consider the following varying coefficient linear model:

$$y_i=f(\boldsymbol{u}_i,\boldsymbol{x}_i)+\sigma(\boldsymbol{u}_i,\boldsymbol{x}_i)\epsilon_i=\beta_0(\boldsymbol{u}_i)+\sum_{j=1}^{d_x}\beta_j(\boldsymbol{u}_i)x_{ij}+\sigma(\boldsymbol{u}_i,\boldsymbol{x}_i)\epsilon_i, \tag{1}$$

where x_{ij} is the jth component of $\boldsymbol{x}_i$, $\beta_j(\cdot)$'s are unknown coefficient functions, $\mathrm{Var}(y_i)=\sigma^2(\boldsymbol{u}_i,\boldsymbol{x}_i)>0$ and ϵ_i's are i.i.d. random variables with mean 0 and variance 1. The varying coefficient model is flexible in that the responses are linearly associated with a set of covariates, but their regression coefficients can vary with smoothing vector $\boldsymbol{u}$.

For the varying coefficient model (1) we now assume that $\beta_j(\boldsymbol{u}_i)$ for $j=0,1,\ldots,d_x$ is nonlinearly related to the smoothing vector $\boldsymbol{u}_i$ such that $\beta_j(\boldsymbol{u}_i)=\boldsymbol{w}_j^t\boldsymbol{\phi}(\boldsymbol{u}_i)+b_j$, where $\boldsymbol{w}_j$ is a corresponding weight vector of dimension d_h to $\boldsymbol{\phi}(\boldsymbol{u}_i)$, and $\boldsymbol{\phi}$ is the nonlinear feature mapping function which maps the input space to the higher dimensional feature space where the dimension d_h is defined in an implicit way. An inner product in feature space has an equivalent kernel in input space, $K(\boldsymbol{u}_i,\boldsymbol{u}_j)=\boldsymbol{\phi}(\boldsymbol{u}_i)^t\boldsymbol{\phi}(\boldsymbol{u}_j)$, provided certain conditions hold [10]. Several choices of the kernel function are possible. Two popular choices of kernel function in practice are Gaussian kernel and polynomial kernel of degree d defined, respectively, as

$$K(\boldsymbol{u}_i,\boldsymbol{u}_j)=\exp\left(-\|\boldsymbol{u}_i-\boldsymbol{u}_j\|^2/2\kappa\right),$$
$$K(\boldsymbol{u}_i,\boldsymbol{u}_j)=(1+\boldsymbol{u}_i^t\boldsymbol{u}_j)^d,\quad i,j=1,\ldots,n,$$

where $\kappa>0$ and d are prespecified kernel parameters. Then, the regression function $f(\cdot)$ of the model (1) can be rewritten as

$$f(\boldsymbol{u}_i,\boldsymbol{x}_i)=\boldsymbol{w}_0^t\boldsymbol{\phi}(\boldsymbol{u}_i)+b_0+\sum_{j=1}^{d_x}x_{ij}\left(\boldsymbol{w}_j^t\boldsymbol{\phi}(\boldsymbol{u}_i)+b_j\right).$$

Using the basic idea of LS-SVR we can define the optimization problem:

$$\min_{\boldsymbol{w}_j,b_j,e}\mathcal{J}=\frac{1}{2}\sum_{j=0}^{d_x}\|\boldsymbol{w}_j\|^2+\frac{\gamma}{2}\sum_{i=1}^{n}e_i^2$$

subject to the equality constraints

$$y_i=\boldsymbol{w}_0^t\boldsymbol{\phi}(\boldsymbol{u}_i)+b_0+\sum_{j=1}^{d_x}x_{ij}\left(\boldsymbol{w}_j^t\boldsymbol{\phi}(\boldsymbol{u}_i)+b_j\right)+e_i,\quad i=1,\ldots,n,$$

where γ is the regularization parameter and e_i's are i.i.d. random variables with mean 0 and $Var(e)<\infty$. We construct the Lagrangian

$$\mathcal{L}=\mathcal{J}-\sum_{i=1}^{n}\alpha_i\left(\boldsymbol{w}_0^t\boldsymbol{\phi}(\boldsymbol{u}_i)+b_0+\sum_{j=1}^{d_x}x_{ij}\left(\boldsymbol{w}_j^t\boldsymbol{\phi}(\boldsymbol{u}_i)+b_j\right)+e_i-y_i\right),$$

where α_i's are the Lagrange multipliers. Then, the Karush–Kuhn–Tucker conditions for optimality are given by

$$\frac{\partial\mathcal{L}}{\partial\boldsymbol{w}_0}=\boldsymbol{0}\rightarrow\boldsymbol{w}_0=\sum_{i=1}^{n}\alpha_i\boldsymbol{\phi}(\boldsymbol{u}_i)$$
$$\frac{\partial\mathcal{L}}{\partial\boldsymbol{w}_j}=\boldsymbol{0}\rightarrow\boldsymbol{w}_j=\sum_{i=1}^{n}\alpha_ix_{ij}\boldsymbol{\phi}(\boldsymbol{u}_i),\quad j=1,\ldots,d_x,$$
$$\frac{\partial\mathcal{L}}{\partial b_0}=0\rightarrow\sum_{i=1}^{n}\alpha_i=0,$$
$$\frac{\partial\mathcal{L}}{\partial b_j}=0\rightarrow\sum_{i=1}^{n}\alpha_ix_{ij}=0,\quad j=1,\ldots,d_x$$
$$\frac{\partial\mathcal{L}}{\partial e_i}=0\rightarrow e_i=\frac{1}{\gamma}\alpha_i,\quad i=1,\ldots,n,$$
$$\frac{\partial\mathcal{L}}{\partial\alpha_i}=0\rightarrow\boldsymbol{w}_0^t\boldsymbol{\phi}(\boldsymbol{u}_i)+b_0+\sum_{j=1}^{d_x}x_{ij}\left(\boldsymbol{w}_j^t\boldsymbol{\phi}(\boldsymbol{u}_i)+b_j\right)+e_i-y_i,\quad i=1,\ldots,n.$$

After eliminating e_i's, $\boldsymbol{w}_0$ and $\boldsymbol{w}_j$'s, we have the optimal values of α_i's, b_0 and b_j's are obtained from the linear equation as follows:

$$\begin{pmatrix}\boldsymbol{K}+\boldsymbol{X}\boldsymbol{X}^t\odot\boldsymbol{K}+\frac{1}{\gamma}\boldsymbol{I}_n & \boldsymbol{1}_n & \boldsymbol{X}\\ \boldsymbol{1}_n^t & 0 & \boldsymbol{0}_{d_x}^t\\ \boldsymbol{X}^t & \boldsymbol{0}_{d_x} & \boldsymbol{O}_{d_x}\end{pmatrix}\begin{pmatrix}\boldsymbol{\alpha}\\ b_0\\ \boldsymbol{b}\end{pmatrix}=\begin{pmatrix}\boldsymbol{y}\\ 0\\ \boldsymbol{0}_{d_x}\end{pmatrix},$$

where $\boldsymbol{X}=(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)^t$, $\boldsymbol{\alpha}=(\alpha_1,\ldots,\alpha_n)^t$, $\boldsymbol{b}=(b_1,\ldots,b_{d_x})^t$, $\boldsymbol{y}=(y_1,\ldots,y_n)^t$, $\boldsymbol{I}_m$ and $\boldsymbol{O}_m$, respectively, denote the identity and zero matrices of dimension m, $\boldsymbol{1}_m$ and $\boldsymbol{0}_m$, respectively, denote the vectors of ones and zeros of dimension m, $\boldsymbol{K}$ is the $n\times n$ kernel matrix with elements $K_{ij}=K(\boldsymbol{u}_i,\boldsymbol{u}_j)$, and $\odot$ denotes a componentwise multiplication.

For a point $(\boldsymbol{u}_o,\boldsymbol{x}_o)$ the VC-LS-SVR method for coefficient functions estimation takes the form:

$$\hat{\beta}_0(\boldsymbol{u}_o)=\sum_{i=1}^{n}K(\boldsymbol{u}_o,\boldsymbol{u}_i)\hat{\alpha}_i+\hat{b}_0,$$
$$\hat{\beta}_j(\boldsymbol{u}_o)=\sum_{i=1}^{n}x_{ij}K(\boldsymbol{u}_o,\boldsymbol{u}_i)\hat{\alpha}_i+\hat{b}_j,\quad j=1,\ldots,d_x,$$

and then for regression function estimation takes the form:

$$\hat{f}(\boldsymbol{u}_o,\boldsymbol{x}_o)=\sum_{i=1}^{n}K(\boldsymbol{u}_o,\boldsymbol{u}_i)\hat{\alpha}_i+\hat{b}_0+\sum_{i=1}^{n}\sum_{j=1}^{d_x}x_{oj}x_{ij}K(\boldsymbol{u}_o,\boldsymbol{u}_i)\hat{\alpha}_i+\sum_{j=1}^{d_x}x_{oj}\hat{b}_j. \tag{2}$$

We remark that $(\boldsymbol{u}_o,\boldsymbol{x}_o)$ could be an observation in the training data set $\mathcal{D}$ or a new observation.

2.2. Model selection

We now consider the model selection problem which determines the appropriate hyperparameters of the proposed VC-LS-SVR method. The functional structure of the VC-LS-SVR method is characterized by hyperparameters such as the regularization parameter γ and the kernel parameter κ or d. To choose the values of hyperparameters of the VC-LS-SVR method we first need to consider the cross validation (CV) function as follows:

$$\mathrm{CV}(\boldsymbol{\lambda})=\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\hat{f}^{(-i)}(\boldsymbol{u}_i,\boldsymbol{x}_i|\boldsymbol{\lambda})\right)^2,$$

where $\boldsymbol{\lambda}$ is the set of hyperparameters, and $\hat{f}^{(-i)}(\boldsymbol{x}_i|\boldsymbol{\lambda})$ is the regression function estimated without ith observation.