Journal of Korean
Data & Information Science Society
2003, Vol. 14, No.2 pp. 337~343

Prediction Intervals for LS-SVM Regression using the Bootstrap[1)]

Jooyong Shim[2)] · Changha Hwang[3)]

Abstract

In this paper we present the prediction interval estimation method using bootstrap method for least squares support vector machine(LS-SVM) regression, which allows us to perform even nonlinear regression by constructing a linear regression function in a high dimensional feature space. The bootstrap method is applied to generate the bootstrap sample for estimation of the covariance of the regression parameters consisting of the optimal bias and Lagrange multipliers. Experimental results are then presented which indicate the performance of this algorithm.

Keywords: Least Squares Support Vector Machine, Bootstrap, Lagrange multiplier.

1. Introductions

Support vector machine(SVM), originally introduced by Vapnik(1995, 1998), solves the weak point of neural network such as the existence of local minima in the area of statistical learning theory and structural risk minimization. SVM solutions are characterized by convex optimization problems. Despite of many successful application of SVM in classification and function estimation problem, SVM requires to solve a quadratic program(QP) problem. QP is to optimize a quadratic function over a polyhedron, defined by linear equations and/or inequalities, which is time memory expensive.

A modified version of SVM in a least squares sense has been proposed for classification in Suykens and Vanderwalle(1999). In LS-SVM concerning

---

1) This work is supported by Catholic University of Daegu grant, 2003.
2) Lecturer, Dept of Statistical Information, Catholic University of Daegu, Kyungbuk, 712-702, Korea.
3) Dept of Statistical Information, Catholic University of Daegu, Kyungbuk, 712-702, Korea.

classification problems, we have regression interpretations and direct links to work in classical statistics. In LS-SVM the solution is given by a linear system instead of a QP problem. Taking account of the fact that the computational complexity increases strongly with the number of training data, LS-SVM can be efficiently estimated using iterative methods. The fact that LS-SVM has explicit primal-dual formulations has a number of advantages.

The problem of prediction interval estimation for SVM regression has been studied recently, Seok et al.(2002) presented a Bayesian approach to estimating the prediction intervals for SVM regression and showed SVM regression achieves better performances than the neural networks and the multivariate adaptive regression splines(MARS) in predicting intervals. In this paper we present how to estimate the prediction intervals using the bootstrap method for LS-SVM regression. The bootstrap method is a computer based method for assigning measures of accuracy to statistical estimates, which generates a large number of bootstrap samples by repeatedly resampling the original data set in random manner to provide informations on the distribution of the statistic of interest. A good introduction can be found in Efron and Tbshirani(1993).

The rest of paper is organized as follows. In Section 2 we give an overview of LS-SVM regression. In Section 3 we present the prediction interval estimation method of nonlinear regression by LS-SVM using bootstrap. In Section 4 we perform the numerical studies with simulated data sets. In Section 5 we give the remarks and conclusions.

2. LS-SVM Regression

Let the training data set D be denoted by $\{(x_i, y_i)\}_{i=1}^n$, with each input $x_i \in R^d$ and the output $y_i \in R$. We consider the case of nonlinear regression. Then, we take the form

$$f(x) = w' \phi(x) + b$$

where the term b is a bias term. Here the feature mapping function $\phi(\cdot) : R^d \rightarrow R^{d_f}$ maps the input space to the higher dimensional feature space where the dimension d_f is defined in an implicit way.

The optimization problem is defined with a regularization parameter γ as

$$\text{Minimize} \quad \frac{1}{2} w' w + \frac{\gamma}{2} \sum_{i=1}^n e_i^2 \tag{1}$$

over { ω, b, e } subject to equality constraints

$$y_i = w' \phi(x_i) + b + e_i , \quad i = 1, \cdots, n.$$

The Lagrangian function can be constructed as

$$L(w, b, e : \alpha) = \frac{1}{2} w' w + \frac{\gamma}{2} \sum_{i=1}^{n} e_i^2 - \sum_{i=1}^{n} \alpha_i (w' \phi(x_i) + b + e_i - y_i) \quad (2)$$

where α_i's are the Lagrange multipliers. The conditions for optimality are given by

$$\frac{\delta L}{\delta w} = 0 \rightarrow w = \sum_{i=1}^{n} \alpha_i \phi(x_i)$$

$$\frac{\delta L}{\delta b} = 0 \rightarrow \sum_{i=1}^{n} \alpha_i = 0$$

$$\frac{\delta L}{\delta e_i} = 0 \rightarrow \alpha_i = \gamma e_i, \quad i = 1, \cdots, n$$

$$\frac{\delta L}{\delta e_i} = 0 \rightarrow w' \phi(x_i) + b + e_i - y_i = 0, \quad i = 1, \cdots, n,$$

with solution

$$\begin{bmatrix} 0 & 1' \\ 1 & \Omega + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix} \quad (3)$$

with $\mathbf{y} = (y_1, \cdots, y_n)'$, $1 = (1, \cdots, 1)'$, $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_n)'$, and $\Omega = \{\Omega_{kl}\}$, where $\Omega_{kl} = \phi(x_k)' \phi(x_l) = K(x_k, x_l)$, $k, l = 1, \cdots, n$, which are obtained from the application of Mercer's conditions(1909). Several choices of the kernel $K(\cdot, \cdot)$ are possible.

Solving the linear equation (3) the optimal bias and Lagrange multipliers, $\hat{b}$ and $\hat{\alpha}_i$'s are obtained, then the optimal regression function for the given x is obtained as

$$\hat{f}(x) = \sum_{i=1}^{n} \hat{\alpha}_i K(x_i, x) + \hat{b} \quad (4)$$

Note that in the nonlinear setting, the optimization problem corresponds to finding the flattest function in the feature space, not in the input space. In fact, SVM has strong advantage that SVM performs particularly well for the nonlinear regression model with several input variables.

3. Prediction Interval Estimation

Assume a nonlinear regression model can be expressed as

$$y_i = f(x_i) + \varepsilon_i = H_i \theta + \varepsilon_i, \quad i = 1, \cdots, n,$$

where $H_i = (1, K(x_i, x_1), \cdots, K(x_i, x_n))$, $\theta = (b, \alpha_1, \cdots, \alpha_n)'$ and ε_i's are unobservable random errors which are independent and identically distributed with zero mean and finite variance σ^2. Here b and α_i's are the bias and the

Lagrange multipliers defined in the equation (2), and $K(\cdot,\cdot)$ is a kernel. θ is estimated from the training data set $\{(x_i, y_i)\}_{i=1}^n$ by LS-SVM regression, which is denoted by $\hat{\theta}$. The residuals are obtained by $\hat{\varepsilon}_i = y_i - H_i\hat{\theta},\ i=1,\cdots,n,$ which leads the estimate of σ^2 as $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n \hat{\varepsilon}_i^2$.

The Bootstrap estimate of the covariance of $\hat{\theta}$ can obtained as follows:

i) Bootstrap data set, $\{(x_i^*, y_i^*)\}_{i=1}^n$, is drawn with replacement from training data set $\{(x_i, y_i)\}_{i=1}^n$.

ii) $\hat{\theta}^*$ is estimated from the bootstrap data set , $\{(x_i^*, y_i^*)\}_{i=1}^n$, by LS-SVM regression.

iii) i) and ii) are repeated B of times to obtain $\hat{\theta}^{*1},\cdots,\hat{\theta}^{*B}$.

iv) $\widehat{COV}(\hat{\theta}) = \frac{1}{B-1}\sum_{b=1}^B(\hat{\theta}^{*b} - \bar{\theta}^*)(\hat{\theta}^{*b} - \bar{\theta}^*)'$,

where $\bar{\theta}^* = \frac{1}{B}\sum_{b=1}^B \hat{\theta}^{*b}$.

Then $(1-\eta)100$ % prediction limits for $Y(x_{new})$ can be obtained as

$$\hat{Y}(x_{new}) \pm z_{\eta/2}\sqrt{H_{new}\widehat{COV}(\hat{\theta})H_{new}' + \hat{\sigma}^2} \qquad (5)$$

where $\hat{Y}(x_{new}) = H_{new}\hat{\theta},\ \ H_{new} = (1, K(x_{new}, x_1),\cdots,K(x_{new}, x_n))$.

4. Numerical Studies

We illustrate the performance of the proposed estimation method through one simulated example. The simulated data set consists of 200 of x generated from a uniform distribution $U(0,1)$ and 200 of y generated from a normal distribution $N(0.5+0.4\times\sin(2\pi x), 0.01)$. The first 100 of data are used for the training data set to estimate θ and the rest of data are used for test data set to estimate the prediction intervals.

For the estimation of $\theta = (b, \alpha_1,\cdots,\alpha_{100})'$ the training data set $\{(x_i, y_i)\}_{i=1}^{100}$, and the radial basis function(RBF) kernel are used, where the RBF kernel is defined as

$$K(x_1, x_2) = \exp(-\frac{1}{2\sigma^2}(x_1 - x_2)'(x_1 - x_2)).$$

The values of γ and σ in RBF kernel are chosen as 500 and 0.2, respectively, by the cross validation of training data set. The 500 bootstrap samples are drawn to estimate the covariance of $\hat{\theta}$.

With the input data of the test data set $\{x_{t_i}\}_{i=1}^{100}$, 95% prediction intervals for each x_{t_i}, $i=1,\cdots,100$, are obtained by the equation (5). The figure 1 shows the predicted target values, $\hat{Y}(x_{t_i})$, and 95% prediction intervals of the target values $Y(x_{t_i})$ by LS-SVM using bootstrap.

Using that the distribution of target values $Y(x_{t_i})$ is known as $N(0.5+0.4\times\sin(2\pi x_{t_i}), 0.01)$ 95% prediction intervals of the target values can be obtained as $0.5+0.4\times\sin(2\pi x_{t_i}) \pm 0.196$. Figure 2 shows the mean target values and 95% prediction intervals of the target values $Y(x_{t_i})$ by the known distribution of target values. By comparing two figures we can see that the proposed estimation provides reasonable estimations of the predicted target values and the corresponding prediction intervals.

Figure 1. 95 % Prediction intervals for the test data set by LS-SVM using bootstrap

Figure 2. 95 % Prediction intervals for the test data set by the known distribution of target values

5. Remarks and Conclusions

Through the example we showed that the proposed algorithm derives the satisfying results, which is attractive approaches to modelling the training data set for the prediction intervals estimation of test data set. In particular, we can use this algorithm successfully even for a linear regression model by using identity feature mapping function, that is, $\phi(x) = x$ which implies the linear kernel such that $K(x_1, x_2) = x_1' x_2$.

In this paper we proposed the prediction interval estimation for LS-SVM regression using the bootstrap method when error terms are independent on the input data. In future work, we intend to devise the prediction interval estimation for standard SVM when error terms are dependent on the input data using other efficient resampling methods.

References

1. Efron, B. and Tibshirani, R. (1993). An Introduction to the Bootstrap, *Chapman Hall,* London.
2. Mercer, J. (1909). Functions of Positive and Negative Type and Their Connection with Theory of Integral Equations, *Philosophical Transactions of Royal Society, A:415-446.*
3. Seok, K., Hwang, C. ,and Cho, D. (2002). Prediction Intervals for Support Vector Machine Regression, *Communications in Statistics: Theory and Methods,* 31, 10, 1887-1898.
4. Suykens, J.A.K. and Vanderwalle, J. (1999). Least Square Support Vector Machine Classifier, *Neural Processing Letters*, 9, 293-300.
5. Vapnik, V. N. (1995). The Nature of Statistical Learning Theory. *Springer,* New York.
6. Vapnik, V. N. (1998). Statistical Learning Theory. *Springer,* New York.

[received date : Dec. 2002, accepted date : Apr. 2003]