The Korean Communications in Statistics
Vol. 13 No. 3, 2006, pp.765-776

Weighted LS-SVM Regression for Right Censored Data

Daehak Kim[1)] and Hyeong Chul Jeong[2)]

Abstract

In this paper we propose an estimation method on the regression model with randomly censored observations of the training data set. The weighted least squares support vector machine regression is applied for the regression function estimation by incorporating the weights assessed upon each observation in the optimization problem. Numerical examples are given to show the performance of the proposed estimation method.

Keywords : Regression model; right censoring; support vector machine.

1. Introduction

The least squares support vector machine(LS-SVM), a modified version of support vector machine(SVM) introduced by Vapnik (1995, 1998) has been proposed for classification and regression by Suykens and Vanderwalle (1999). In LS-SVM the solution is given by a linear system instead of a quadratic program problem. Taking account of the fact that the computational complexity increases strongly as the number of training data becomes larger, LS-SVM regression can be estimated efficiently for the huge data set by using iterative methods.

The accelerated failure time model(AFT) and the least squares method to accommodate the censored data seem appealing since they are familiar and well understood. Koul *et al.* (1981) gave a simple least squares type estimation procedure in the censored regression model with the weighted observations and also showed the consistency and asymptotic normality of the estimator. Zhou (1992) proposed an M-estimator of the regression parameter based on the censored data using the similar weighting scheme as Koul *et al.* (1981) proposed. In this paper we consider the estimators of regression parameters and the fitted regression function by weighted LS-SVM based on the right censored

1) Professor, Department of Statistical Information, Catholic University of Daegu, Kyungbuk 712-702, Korea.
Correspondence : dhkim@cu.ac.kr
2) Assistant Professor, Department of Applied Statistics, The University of Suwon, Suwon, Kyunggi, Korea.

observations of the training data set. The similar weighting scheme as Zhou (1992) used and the squared error loss function are included in the optimization problem of weighted LS-SVM. In section 2, we give an overview of LS-SVM regression. In section 3 we suggest an estimation method on the regression model with randomly right censored data by weighted LS-SVM. Numerical studies with Stanford heart transplant data set were performed in section 4. Finally we give a concluding remarks in section 5.

2. Least Squares Support Vector Machines

Let the training data set be denoted by $\{\mathbf{x}_i, y_i\}_{i=1}^{n}$, with each input $\mathbf{x}_i \in \mathbf{R}^d$ and the output $y_i \in \mathbf{R}$. In this section we give an overview of LS-SVM regression for linear and nonlinear cases, respectively.

2.1 Linear regression

For the case of well known linear regression, we can assume the form of unknown regression function f for given input vector $\mathbf{x}$ by

$$f(\mathbf{x}) = \mathbf{w}'\mathbf{x} + \mathrm{b} \tag{1}$$

where b is a bias term and $\mathbf{w}$ is an appropriate weight vector. LS-SVM approach to minimizing the guaranteed risk bound for linear model leads to the optimization problem defined with a regularization parameter γ as

$$\min_{\mathbf{w}} \ \frac{1}{2}\mathbf{w}'\mathbf{w} + \frac{\gamma}{2}\sum_{i=1}^{n} e_i^2 \tag{2}$$

over $(\mathbf{w}, b, \mathbf{e})$ subject to equality constraints

$$y_i - \mathbf{w}'\mathbf{x}_i - b = e_i, \ i = 1, \cdots, n \tag{3}$$

where $\mathbf{e} = (e_1, \cdots, e_n)$. The Lagrangian function can be constructed as

$$L(\mathbf{w}, b, e : \alpha) = \frac{1}{2}\mathbf{w}\mathbf{w}' + \frac{\gamma}{2}\sum_{i=1}^{n} e_i^2 - \sum_{i=1}^{n}\alpha_i(\mathbf{w}'\mathbf{x}_i + b + e_i - y_i) \tag{4}$$

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

$$\frac{\partial L}{\partial \mathbf{w}} = 0 \rightarrow w = \sum_{i=1}^{n}\alpha_i \mathbf{x}_i$$

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

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

$$\frac{\partial L}{\partial \alpha_i} = 0 \rightarrow \mathbf{w}'\mathbf{x}_i + b + e_i - y_i = 0, \ i = 1, \cdots, n,$$

with solution

$$\begin{bmatrix} o & 1' \\ 1 & \Omega_L+\gamma^{-1}\mathbf{I} \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix} \tag{5}$$

with $\mathbf{y}=(y_1,\cdots,y_n)', \mathbf{1}=(1,\cdots,1)', \alpha=(\alpha_1,\cdots,\alpha_n)'$ and $\Omega_L=\{\mathbf{x}'_k\mathbf{x}_l\},\ k,l=1,\cdots,n.$ Solving the linear equation (5) the estimators of the optimal bias and Lagrange multipliers, $\hat{b}$ and $\hat{\alpha}_i$'s can be obtained. And then the estimator of the regression parameters can be obtained as

$$\hat{f}=\sum_{i=1}^{n}\hat{\alpha}_i\mathbf{x}'_i\mathbf{x}+\hat{b} \tag{6}$$

2.2 Nonlinear Regression

To allow for the case of nonlinear regression, the input vectors are nonlinearly transformed into a potentially higher dimensional feature space by a nonlinear mapping function $\phi(\bullet)$ and then a linear regression is performed there. Nonlinear regression function can be written as

$$f(\mathbf{x})=\mathbf{w}'\phi(\mathbf{x})+b \tag{7}$$

where b is a bias term, $\mathbf{w}$ is an appropriate weight vector and $\phi(\bullet)$ is a nonlinear mapping function. LS-SVM approach for nonlinear model leads to the optimization problem defined with a regularization parameter γ as

$$\min_{\mathbf{w}}\ \frac{1}{2}\mathbf{w}'\mathbf{w}+\frac{\gamma}{2}\sum_{i=1}^{n}e_i^2 \tag{8}$$

over $(\mathbf{w},b,\mathbf{e})$ subject to equality constraints

$$y_i=\mathbf{w}'\phi(\mathbf{x}_i)-b=e_i,\quad i=1,\cdots,n. \tag{9}$$

The Lagrangian function can be constructed as

$$L(\mathbf{w},b,e:\alpha)=\frac{1}{2}\mathbf{w}'\mathbf{w}+\frac{\gamma}{2}\sum_{i=1}^{n}e_i^2-\sum_{i=1}^{n}\alpha_i(\mathbf{w}'\phi(\mathbf{x}_i)+b+e_i-y_i) \tag{10}$$

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

$$\frac{\partial L}{\partial \mathbf{w}}=0\rightarrow\mathbf{w}=\sum_{i=1}^{n}\alpha_i\phi(\mathbf{x}_i)$$

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

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

$$\frac{\partial L}{\partial \alpha_i}=0\rightarrow\mathbf{w}'\phi(\mathbf{x}_i)+b+e_i-y_i=0,\ i=1,\cdots,n,$$

Thus for the case of nonlinear regression, the estimators of the optimal bias and Lagrange multipliers can be obtained by solving the linear equation

$$\begin{bmatrix} 0 & 1' \\ 1 & \Omega_N + \gamma^{-1}\mathbf{I} \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix} \tag{11}$$

with $\mathbf{y} = (y_1, \cdots, y_n)'$, $1 = (1, \cdots, 1)'$, $\alpha = (\alpha_1, \cdots, \alpha_n)'$ and $\Omega_N = \{K_{kl}\}, k, l = 1, 2, \cdots, n$, where

$$K_{kl} = \phi(\mathbf{x}_k)'\phi(\mathbf{x}_l).$$

For this nonlinear regression, solution of (11) requires the computations of dot products $\phi(\mathbf{x}_k)'\phi(\mathbf{x}_l)$, $k, l = 1, \cdots, n$, in a potentially higher dimensional feature space. Under certain conditions(Mercer, 1909), these demanding computations can be reduced significantly by introducing a kernel function such that

$$\phi(\mathbf{x}_k)'\phi(\mathbf{x}_l) = K(\mathbf{x}_k, \mathbf{x}_l).$$

Several choices of kernel functions are possible. RBF(Radial Basis Function) is the most frequently used kernel function. In the nonlinear case we can not obtain the estimators of regression parameters corresponding to the nonlinear feature mapping function of $\mathbf{x}$ explicitly, but the optimal nonlinear regression function for the given $\mathbf{x}$ can be obtained as

$$\hat{f}(\mathbf{x}) = \sum_{i=1}^{n} \hat{\alpha}_i K(\mathbf{x}_i, \mathbf{x}) + \hat{b}. \tag{12}$$

The linear regression model (1) can be regarded as a special case of nonlinear regression model (7). By using an identity feature mapping function $\phi(\cdot)$ in nonlinear regression model, that is, $K(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{x}_1'\mathbf{x}_2$, it reduces to linear regression model.

3. Regression with Censored Data by Weighted LS-SVM

In this section we suggest an estimation method on the regression model with randomly right censored data by weighted LS-SVM. For the suggestion, we consider the censored linear regression model first and then extend the result of censored linear regression model to censored nonlinear regression model.

Consider the censored linear regression model for the response variables T_i's,

$$T_i = \beta'\mathbf{x}_i + b + \epsilon_i, \quad i = 1, \cdots, n,$$

where $(\beta', b)'$ is the regression parameter vector of the model and ϵ_i's are unobservable errors assumed to be independent with zero means and bounded variances. Let C_i's be the censoring variables assumed to be independent and identically distributed having a cumulative distribution function $G(y) = P(C_i \leq y)$. The parameter vector of interest is $(\beta', b)'$ and T_i is not observed but

$$\delta_i = I_{(T_i < c_i)} \text{ and } y_i = \min(T_i, C_i),$$

where $I_{(\cdot)}$ denotes the indicator function. In most practical cases $G(\cdot)$ is not known and needs to be estimated by the Kaplan-Meier estimator (1958) or its variation, $\hat{G}(\bullet)$. The problem considered here is that of the estimation of $(\beta',b)'$ based on $(\delta_1,y_1,\mathbf{x}_1),\cdots,(\delta_n,y_n,\mathbf{x}_n)$. Koul *et al.* (1981) defined a new observable response Y_i^* with weights $\hat{\zeta}_i$ as

$$Y_i^* = \hat{\zeta}_i y_i \ , \ \hat{\zeta}_i = \frac{\delta_i}{1-\hat{G}(y_i)},$$

and showed Y_i^* has the same mean as T_i and thus follows the same linear model as T_i does. Koul *et al.* (1981) used the variation of Kaplan-Meier estimator of $G(\bullet)$,

$$\hat{G}(t) = 1 - \prod_{i:y(i)\leq t}(1-\frac{d(i)}{1+n(i)})^{1-\delta_i},$$

where $n(y)$ is the number of alives at y^- and $d(y)$ is the number of deaths at y where y^- means just before y. The estimator of $(\beta',b)'$ proposed by Koul *et al.* (1981) is obtained from

$$(\hat{\beta},\hat{b}) = \arg\min_{(\beta,b)}\sum_{i=1}^{n}(Y_i^* - \beta'\mathbf{x}_i - b)^2.$$

Zhou (1992) proposed an M-estimator of the regression parameter with a general loss function $\rho(\bullet)$using the weightsζ_i,

$$(\hat{\beta},\hat{b}) = \arg\min_{(\beta,b)}\sum_{i=1}^{n}\hat{\zeta}_i\rho(Y_i - \beta'\mathbf{x}_i - b).$$

We use the left continuous version of Kaplan-Meier estimator as Zhou (1992) used,

$$\hat{G}(t) = 1 - \prod_{i:y(i)<t}(1-\frac{d(i)}{n(i)})^{1-\delta_i}, \tag{13}$$

And we obtain the estimator of weights as follows:

$$\hat{\zeta}_i = \frac{\delta_i}{1-\hat{G}(y_i)}, \quad i=1,\cdots,n, \tag{14}$$

δ_i is set 1 for the maximum order statistic $y_{(n)}$ so that no weight is lost even if the largest observation is censored. We apply the weighting scheme of Koul *et al.* (1981) to (2) with squared error loss function. Then the optimal problem of the weighted LS-SVM can be constructed as

$$\min\frac{1}{2}\beta'\beta + \frac{\gamma}{2}\sum_{i=1}^{n}\hat{\zeta}_i e_i^2, \tag{15}$$

with the equality constraints as

$$y_i - \beta'\mathbf{x}_i - b = e_i, \quad i=1,\cdots,n.$$

Thus the estimators of the optimal bias and Lagrange multipliers, $\hat{b}$ and $\hat{\alpha}_i$'s, can be obtained from the linear equation

$$\begin{bmatrix} 0 & 1' \\ 1 & \Omega+\gamma^{-1}\mathrm{Diag}\{\hat{\zeta}_i\}^{-1} \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix} \tag{16}$$

where

$$\Omega = \{\mathbf{x}'_i \mathbf{x}_j\},\ \hat{\zeta}_i = \frac{\delta_i}{1-\hat{G}(y_i)},\ i,j=1,\cdots,n.$$

Solving the above linear equation the optimal bias and Lagrange multipliers, $\hat{b}$ and $\hat{\alpha}_i$'s can be obtained. And then the estimator of the regression parameters are obtained as

$$\hat{\beta} = \sum_{i=1}^{n} \hat{\alpha}_i \mathbf{x}_i \tag{17}$$

For the censored nonlinear regression using kernel function and feature mapping function mentioned in section 2.2, the estimators of the optimal bias and Lagrange multipliers, $\hat{b}$ and $\hat{\alpha}_i$'s, are obtained by solving the following linear equation

$$\begin{bmatrix} 0 & 1' \\ 1 & \Omega+\gamma^{-1}\mathrm{Diag}\{\hat{\zeta}_i\}^{-1} \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix} \tag{18}$$

where

$$\Omega = \{K(\mathbf{x}'_i, \mathbf{x}_j)\},\ \hat{\zeta}_i = \frac{\delta_i}{1-\hat{G}(y_i)},\ i,j=1,\cdots,n.$$

Then the optimal nonlinear regression function for given $\mathbf{x}$ is predicted as

$$\hat{f}(\mathbf{x}) = \sum_{i=1}^{n} \hat{\alpha}_i K(\mathbf{x}_i, \mathbf{x}) + \hat{b}. \tag{19}$$

4. Numerical Studies

We illustrate the performance of the LS-SVM for estimation of the survival pattern of Stanford heart transplant data set(Miller and Halpern, 1982). Among 152 patients with complete record who survived at least 10 days were 55 censored observations. We first consider the quadratic age model which has been considered by Miller and Halpern (1982) in an attempt to achieve better fit of data rather than the multiple regression model with age and T5 mismatch score. Let T_i be the base 10 logarithm of the survival time of the ith patient. To examine the age effect, we use the regression model as follows:

$$T_i = \beta_1 x_i + \beta_2 x_i^2 + b, \tag{20}$$

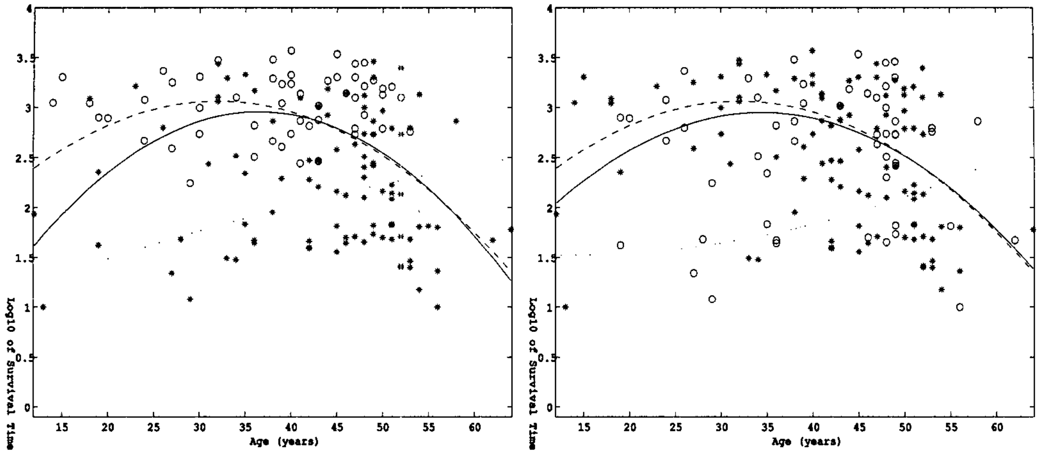
where x_i is the age at the first transplant time, $i=1,\cdots,152$. We apply (13) and (14) to the data to form the estimates of weights. After the estimates of the weights are formed, the estimators of regression parameters can obtained by (16) and (17). The value of γ in (16) was chosen as 31 by 10 fold cross-validation with uncensored observations. To obtain better results, the entire data set can be divided into four strata according to age: < 30, 30 to 40^-, 40 to 50^-, and ≥ 50. The strata contain 30, 23, 66 and 38 patients, respectively. This stratification is known to provide more reliable result and also used by Leurgans (1987), Fygenson and Zhou (1992). We apply (13) and (14) to each stratum to form the estimates of weights. For the Koul *et al.* (1981) estimator, the variation of the Kaplan-Meier estimator of the censoring distribution is employed as they proposed (1981). <Table 1> gives the estimators for the model with covariates age and age squared on the stanford transplant data with 152 patients. Buckley-James estimators in <Table 1> are from Zhou (1992, Table 1)

<Table 1> Estimated regression parameters for $\log_{10}$ survival times on age and age squared.

Method	Intercept	Age	Age squared
Buckley-James	1.353	0.1069	-0.0016700
weighted LS-SVM	-0.0207403	0.1632452	-0.0022371
Zhou with $\rho(t)=t^2$	-0.0207599	0.1632584	-0.0022373
Koul *et al.*	0.8431440	0.0350696	-0.0001454
weighted LS-SVM(stratified)	0.8188321	0.1236407	-0.0017934
Zhou with $\rho(t)=t^2$(stratfied)	0.8186594	0.1236501	-0.0017935
Koul *et al.*(stratified)	1.5989913	-0.0115523	0.0004345

Both weighted LS-SVM estimator(proposed) and Zhou estimator give almost same values of regression parameters regardless of stratification.

<Figure 1> represents $\log_{10}$ survival times versus age for 152 Stanford heart transplant patients under the regression model (20) without stratification(left) and with stratification(right). In <Figure 1>, patients denoted by * are uncensored and those by o are censored. Solid line is the weighted LS-SVM estimator, dashed line is James-Buckley estimator and dotted line is Koul *et al.* estimator. The plots of fitted values of regression functions show the estimators of weighted LS-SVM agree quite well to the Buckley-James estimators for age 40 to 65, which was known well fitted on this data set(Miller and Halpern, 1982). The Koul *et al.* estimator contradicts the fact that younger patients had a better survival after transplant than older patients(Miller and Halpern, 1982), stratification does not resolve this discrepancy remarkably. While weighted LS-SVM estimator is less sensitive than Koul *et al.* estimator against the stratification.



<Figure 1> Scatter plot of $\log_{10}$ survival times versus age with model (20).
(left = without stratification, right = with stratification)

Now we consider the multiple regression model including T5 mismatch score which was deleted from the quadratic age model due to its nonsignificance(Miller and Halpern, 1982) as follows:

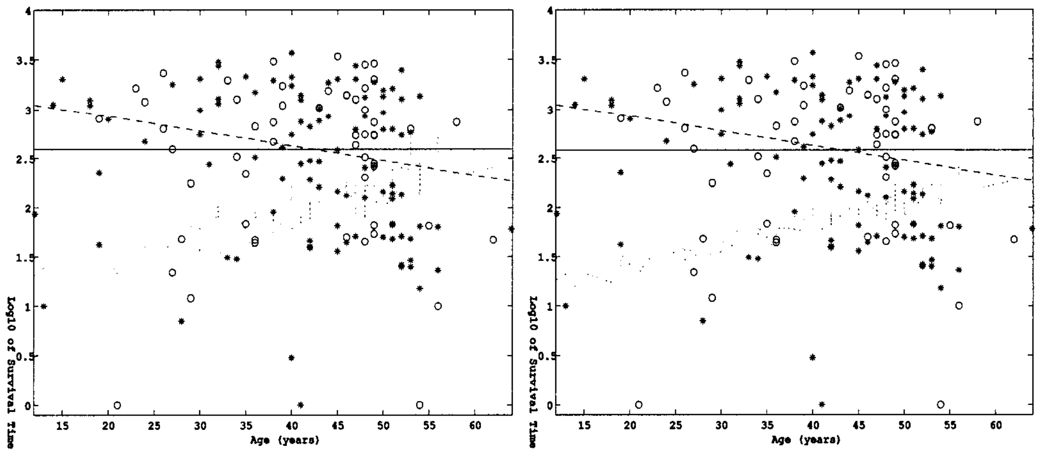
$$T_i = \beta_1 x_{1i} + \beta_2 x_{2i} + b, \tag{21}$$

where x_{1i} is the age at the first transplant time and x_{2i} is the T5 mismatch score, $i = 1, \cdots, 157$. Here each of 157 patients has complete record of age and T5 mismatch score. <Table 2> gives the estimators for the model with covariates age and T5 mismatch score on the Stanford transplant data with 157 patients. Buckley-James estimator and weighted LS-SVM estimator agree in a sense that the age effect on survival is relatively larger than that of T5 mismatch score, but other estimators show it reversely.

<Figure 2> represents $\log_{10}$ survival times versus age for 157 Stanford heart transplant patients under the regression model (21) without stratification and with stratification, respectively. In <Figure 2> we can see the Koul *et al.* estimator contradicts the fact that younger patients had a better survival after transplant than older patients(Miller and Halpern, 1982). Buckley-James estimator exceeds most of data at the low values of age, this implies the nonlinear regression model may describe the age effect better than the linear regression model.

<Table 2> Estimated regression parameters for $\log_{10}$ survival times on age and T5 mismatch score.

Method	Intercept	Age	T5 mismatch score
Buckley-James	0.35	-0.015	-0.003
weighted LS-SVM	2.5912889	-0.0000092	-0.0000015
Zhou with $\rho(t) = t^2$	3.0391097	-0.0043546	-0.2544753
Koul *et al.*	0.7144260	0.0237964	0.2475615
weighted LS-SVM(stratified)	2.5809569	-0.0000176	-0.0000012
Zhou with $\rho(t) = t^2$(stratfied)	3.1300610	-0.0082401	-0.1954170
Koul *et al.*(stratified)	0.8204691	0.0211247	0.1534609

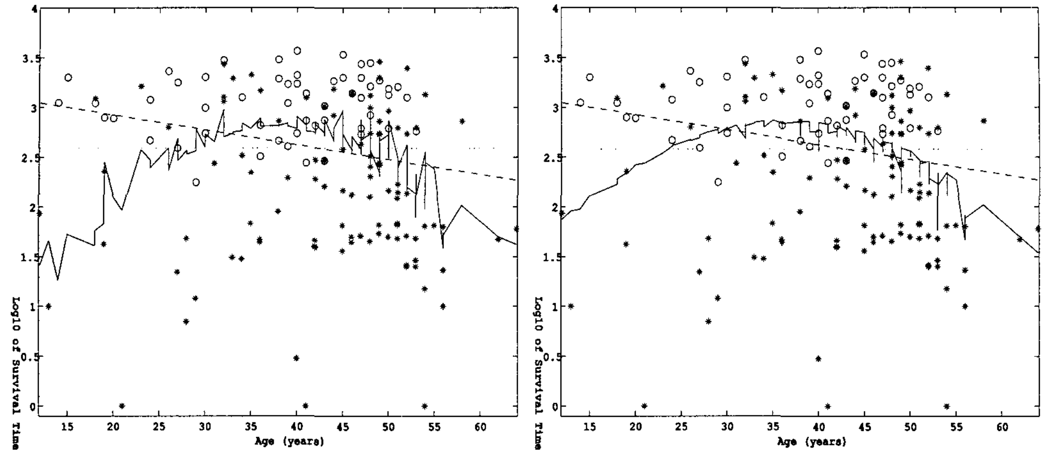


<Figure 2> Scatter plot of $\log_{10}$ survival times versus age with model (21).
(left = without stratification, right = with stratification)

We now consider the nonlinear regression model as follows:

$$T_i = \beta'\phi(\mathbf{x}_i) + b, \tag{22}$$

where $\mathbf{x}_i = (x_{1i}, x_{2i})'$ with x_{1i} the age at the first transplant time and x_{2i} the T5 mismatch score, $i = 1, \cdots, 157$. $\phi(\bullet)$ is the nonlinear feature mapping function such that $\phi(\mathbf{x}_k)'\phi(\mathbf{x}_l) = K(\mathbf{x}_k, \mathbf{x}_l)$, where $K(\bullet, \bullet)$ is the kernel function defined in section 2.2. For this data set, we use the polynomial kernel function $K(\mathbf{x}_k, \mathbf{x}_l) = (1 + \mathbf{x}_k'\mathbf{x}_l)^3$. The value of γ is chosen as 0.0015 by the 10 fold cross-validation method with uncensored observations. <Figure 3> represents $\log_{10}$ survival times versus age for 157 Stanford heart transplant patients under the regression model (22) without stratification and with stratification, respectively. In <Figure 3> we can see the nonlinear LS-SVM estimator gives the similar age effect pattern which appeared in the regression model (20).



<Figure 3> Scatter plot of $\log_{10}$ survival times versus age with model (22). (left = without stratification, right = with stratification)

Now consider the censored nonlinear regression model for the response variables T_i's of the form,

$$T_i = f(x_i) + \epsilon_i, \quad i = 1, \cdots, n. \tag{23}$$

For the data set, 200 of x's are generated from a uniform distribution $U(0,1)$ and 200 of (t,c)'s are generated from the function $0.5 + \sin(0.75\pi x)$. Errors were generated from the double exponential distributions with scale parameter 3. Censoring proportion cc is also chosen as 10%, 25%, and 50% respectively. The radial basis function(RBF) kernel is used for the numerical studies, which is defined as

$$K(x_k, x_l) = \exp[-\frac{1}{2_{\sigma^2}}(x_k - x_l)^2].$$

The values of γ and σ in RBF kernel are chosen by the 10 fold cross-validation method with uncensored observations in each generated data set. Solving the linear equation (18) with the data set, the estimators of the optimal Lagrange multipliers and bias, $\hat{\alpha}_i$'s and $\hat{b}$, can be obtained. Then by the equation (19) the fitted regression function is obtained.

To measure the performance of estimating regression function, we employ the fraction of variance unexplained(FVU), which is given by

$$FVU = \frac{E(\hat{f}(x_i) - f(x_i))^2}{E(f(x_i) - \bar{f}(x))^2},$$

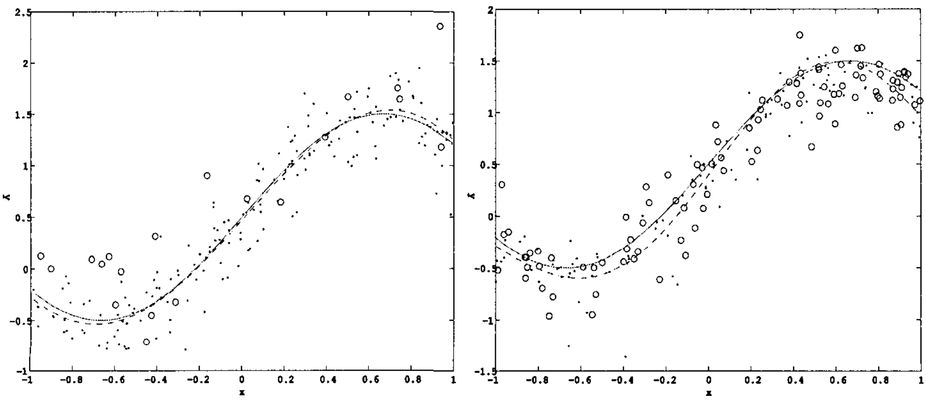
where $\hat{f}(\mathbf{x}_i)$ is the fitted value of the function for a given $\mathbf{x}_i$, $f(\mathbf{x}_i)$ is the true value of the function for a given $\mathbf{x}_i$ and $\bar{f}(\mathbf{x})$ is the average of true values of the function for $\mathbf{x}_1, \cdots \mathbf{x}_n$. Note that the *FVU* is the mean squared error for the estimate $\hat{f}(\mathbf{x})$ scaled by the variance of the true function $f(\mathbf{x})$. We evaluate the *FVU* by replacing the expectations with the average over a set of 200 test set values. With optimal Lagrange multipliers and bias obtained from the training data set, we have the fitted regression functions for each of 100 data sets, thus 100 *FVU*s from 100 data sets.

<Table 3> shows the averages and the root mean squared errors(RMSE) of the 100 *FVU*s obtained by weighted LS-SVM for the fitted regression functions according to various censoring proportions.

<Table 3> The average and RMSE of 100 *FVU*s according to various censoring proportions

	10%	25%	50%
Average	0.0053	0.0111	0.0289
RMSE	0.0037	0.0059	0.0176

The <Figure 4> shows scatter plots of response variables versus explanatory variables for 200 data and estimated regression lines with 10% and 50% censoring proportion, respectively. Data points denoted by * are uncensored and those by o are censored. Solid line is the true regression function, dashed line is the fitted regression function by weighted LS-SVM. The fitted values look close to the true regression functions in this nonlinear model as in linear model for x's from the data set even though more than 25% censoring proportion. For the case of nonlinear regression model, it is hard to find a regression method on the censored data set to compare with the proposed method.



<Figure 4> Scatter plots and estimated regression lines.
(left = 10% censoring proportion, right = 50% censoring proportion)

5. Concluding Remarks

Through the numerical studies, we showed that the proposed method using weighted LS-SVM provides a satisfying solutions to the right censored linear regression model and the censored nonlinear regression model, respectively. Particularly for the censored nonlinear regression model, the proposed method can be used without heavy computations and provides a satisfying result. In future work, we consider to devise algorithms for predicting intervals of regression function based on the training data set which might be randomly right censored, by using weighted LS-SVM or the other efficient machine learning methods.

References

[1] Fygenson, M and Zhou, M. (1992). Modifying the Koul, Susarla and Van Ryzin estimator for linear regression models with right censoring. *Statistics & Probability Letters,* Vol. 13, 295-299.

[2] Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of American Statistical Association*, Vol. 66, 484-491.

[3] Koul, H., Susarla, V. and Van Ryzin J. (1981). Regression Analysis with Randomly Right Censored Data. *The Annals of Statistics*, Vol. 9, 1276-1288.

[4] Leurgans, S. (1987). Linear models, random censoring and synthetic data. *Biometrika,* Vol. 74, 301-9.

[5] Mercer, J. (1909). Functions of Positive and Negative Type and Their Connection with Theory of Integral Equations. *Philosophical Tran -sactions of Royal Society*, A, 415-446.

[6] Miller, R. and Halpern, J. (1982). Regression with Censored Data. *Biometrika*, Vol. 69, 521-531.

[7] Suykens, J.A.K. and Vanderwalle, J. (1999). Least Square Support Vector Machine Classifier. *Neural Processing Letters*, Vol. 9, 293-300.

[8] Vapnik, V.N. (1995). *The Nature of Statistical Learning Theory.* Springer, New York.

[9] Vapnik, V.N. (1998). *Statistical Learning Theory*, Springer, New York.

[10] Zhou, M (1992). M-estimation in Censored Linear Models. *Biometrika*, Vol. 79, 837-841.

[Received September 2006, Accepted October 2006]