Journal of the Korean Data & Information Science Society
2012, **23**(2), 393–401

한국데이터정보과학회지

Two-step LS-SVR for censored regression†

Jongsig Bae[1] · Changha Hwang[2] · Jooyong Shim[3]

[1]Department of Mathematics, Sungkyunkwan University
[2]Department of Statistics, Dankook University
[3]Department of Data Science, Inje University

Received 30 January 2012, revised 1 March 2012, accepted 6 March 2012

Abstract

This paper deals with the estimations of the least squares support vector regression when the responses are subject to randomly right censoring. The estimation is performed via two steps - the ordinary least squares support vector regression and the least squares support vector regression with censored data. We use the empirical fact that the estimated regression functions subject to randomly right censoring are close to the true regression functions than the observed failure times subject to randomly right censoring. The hyper-parameters of model which affect the performance of the proposed procedure are selected by a generalized cross validation function. Experimental results are then presented which indicate the performance of the proposed procedure.

Keywords: Censored regression, generalized cross validation function, Kaplan-Meier estimator, kernel function, least squares support vector machine, randomly right censoring.

1. Introduction

The accelerated failure time model and the least squares method to accommodate the censored data seem appealing since they are familiar and well understood. Miller (1976) proposed a simple estimation in censored regression model by applying the weighted least square method. Koul *et al.* (1981) proposed a censored regression model using the weighted observations. Zhou (1992) proposed the M-estimators of regression parameter with the weights suggested by Koul *et al.* (1981). Yang (1999) proposed a censored median regression model as an alternative to the mean regression model for examining the covariate effect with the data subject to randomly right censoring and showed that the estimators are consistent

† This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2011-0027344).
[1] Professor, Department of Mathematics, Sungkyunkwan University, Suwon 440-746, Korea.
[2] Professor, Department of Statistics, Dankook University, 126, Jukjeon-dong, Suji-gu, Yongin-si, Gyeonggi-do 448-701, Korea.
[3] Corresponding author: Adjunct professor, Institute of Statistical Information, Department of Data Science, Inje University, Obang-Dong, Kimhae 621-749, Korea. E-mail: ds1631@daum.net

and asymptotically normally distributed. Heuchenne and Keilegom (2005) proposed an estimation procedure which extends the least squares procedures for nonlinear regression with censored data.

Support vector machine (SVM) and least squares SVM (LS-SVM) have been very successful in classification and regression problems (Vapnik, 1998; Suykens and Vanderwalle, 1999). Applications of SVM and LS-SVM can be found in Shim (2005), Shim and Lee (2009), Cho *et al.* (2010), Hwang and Shim (2010). In this paper we use LS-SVR and LS-SVR for censored data (LS-SVRc) using a weighting system formed with the Kaplan-Meier estimator (Kaplan and Meier, 1958) of the censoring distribution. The use of the Kaplan-Meier weights to account for censoring has been first proposed by Stute (1993). Shim *et al.* (2011) proposed a semiparametric LS-SVR to consider situations where the functional form of the effect of one or more covariates is unknown in censored data. The existing nonlinear estimation methods of the AFT model have not been widely used in practice, mainly due to their complexity even when the number of covariates is relatively small (Jin *et al.*, 2003). In contrast, the proposed method can be easily applied to the analysis of censored data with high dimensional covariates, since its estimation procedure is done by solving simple linear equation systems.

In this paper we estimate the regression function by LS-SVR for censored regression via two steps - the ordinary LS-SVR and LS-SVRc, where the weighted squared error loss function are included in the optimization problem of LS-SVR. It is well known that the estimation performance of LS-SVR and LS-SVRc are affected the hyper-parameters. The rest of this paper is organized as follows. In Section 2 we give a brief review of LS-SVR. In Section 3 we present the estimation procedure of two-step LS-SVR for AFT model and GCV function for selecting hyper- parameters. In Section 4 we perform the numerical studies through examples. In Section 5 we give the concluding remarks.

2. Least squares support vector for regression

Let the data set be denoted by $\{\boldsymbol{x}_i, y_i\}_{i=1}^n$, with each input vector $\boldsymbol{x}_i \in R^d$ and the response $y_i \in R$ which is the output corresponding to $\boldsymbol{x}_i$. For kernel ridge regression, we can assume the functional form of unknown regression function f for given input vector $\boldsymbol{x}$ by $f(\boldsymbol{x}) = \boldsymbol{\omega}'\phi(\boldsymbol{x}) + b$ where $\boldsymbol{\omega}$ is an appropriate weight vector. 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 C as

$$min \frac{1}{2}\boldsymbol{\omega}'\boldsymbol{\omega} + \frac{C}{2}\sum_{i=1}^n e_i^2 \tag{2.1}$$

over $\{\boldsymbol{\omega}, \boldsymbol{e}\}$ subject to equality constraints,

$$e_i = y_i - \boldsymbol{\omega}'\phi(\boldsymbol{x}_i) - b, \ i = 1, \cdots, n.$$

The Lagrangian function can be constructed as

$$L(\boldsymbol{\omega}, \boldsymbol{e}, b : \boldsymbol{\alpha}) = \frac{1}{2}\boldsymbol{\omega}'\boldsymbol{\omega} + \frac{C}{2}\sum_{i=1}^n e_i^2 - \sum_{i=1}^n \alpha_i(e_i - y_i + \boldsymbol{\omega}'\phi(\boldsymbol{x}_i) + b), \tag{2.2}$$

where α_i's are the Lagrange multipliers. The Karush-Kuhn-Tucker (Smola and Scholkopf, 1998) conditions for optimality are given by

$$\begin{aligned}
&\frac{\partial L}{\partial \boldsymbol{\omega}} = \mathbf{0} \ \rightarrow \boldsymbol{\omega} = \sum_{i=1}^{n} \alpha_i \phi(\boldsymbol{x}_i) \\
&\frac{\partial L}{\partial b} = 0 \ \rightarrow \sum_{i=1}^{n} \alpha_i = 0 \\
&\frac{\partial L}{\partial e_i} = 0 \ \rightarrow \alpha_i = Ce_i, \ i = 1, \cdots, n \\
&\frac{\partial L}{\partial \alpha_i} = 0 \ \rightarrow \boldsymbol{\omega}' \phi(\boldsymbol{x}_i) + e_i - y_i = 0, \ i = 1, \cdots, n,
\end{aligned}$$

leading to the solution

$$\begin{pmatrix} \boldsymbol{\alpha} \\ b \end{pmatrix} = \begin{pmatrix} K + I/C & \mathbf{1} \\ \mathbf{1}' & 0 \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{y} \\ 0 \end{pmatrix} \tag{2.3}$$

with $\boldsymbol{y} = (y_1, \cdots, y_n)'$, $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_n)'$, and $K = \{K_{kl}\}$ where $K_{kl} = \phi(\boldsymbol{x}_k)'\phi(\boldsymbol{x}_l)$, $k,\ l = 1, \cdots, n$, which are obtained from the application of Mercer's conditions (1909). Several choices of the kernel $K(\cdot, \cdot)$ functions are possible (Gunn, 1998). From (2.3) the fitted regression function is obtained as

$$\widehat{f}(\boldsymbol{x}) = K\boldsymbol{\alpha} + b. \tag{2.4}$$

3. Two-step LS-SVR in censored regression

To explain LS-SVR with censored data, we put t_i to be the response variables corresponding to vector, $\boldsymbol{x}_i$ or transformation on it, where $i = 1, 2, \cdots, n$. Let $m(\boldsymbol{x}_i)$ be the regression function of the response variable given $\boldsymbol{x}_i$. We assume that $m(\boldsymbol{x}_i)$ is related to the vector of covariates $\boldsymbol{x}_i$ in a form as

$$m(\boldsymbol{x}_i) = \boldsymbol{\omega}' \phi(\boldsymbol{x}_i) + b \text{ for } i = 1, 2, \cdots, n, \tag{3.1}$$

where b is the bias, $\phi(\boldsymbol{x}_i)$ is a nonlinear feature mapping function.

In fact we can not observe t_i's but the observed variable, $y_i = \min(t_i, c_i)$ and $\delta_i = I(t_i \leq c_i)$, where $I(\cdot)$ denotes the indicator function and c_i is the censoring variable corresponding to x_i for $i = 1, 2, \cdots, n$. c_i's are assumed to be independently distributed with unknown survival distribution functions.

In most practical cases survival distribution function of c_i's, G, is not known and needs to be estimated by the Kaplan-Meier (1958) estimator or its variation. The problem considered here is that of the estimation of $m(\boldsymbol{x}_i)$ based on $(\delta_1, y_1, \boldsymbol{x}_1), \cdots, (\delta_n, y_n, \boldsymbol{x}_n)$. Koul *et al.* (1981) defined new observable responses y_i^* as $y_i^* = u_i y_i$ with

$$u_i = \frac{\delta_i}{G(y_i)}, \tag{3.2}$$

and showed y_i^* has the same mean as t_i and thus follows the same linear model as t_i does. Here, $\widehat{G}$, the Kaplan-Meier estimates (Kaplan and Meir, 1958) of survival distribution function G of c_i's can be obtained as,

$$\widehat{G}(y) = \begin{cases} \prod_{i:y_{(i)}\le y} \left(\frac{n-i}{n-i+1}\right)^{1-\delta_{(i)}}, & \text{if } y \le y_n \\ 0, & otherwise \end{cases}$$

where $(y_{(i)}, \delta_{(i)})$ is (y_i, δ_i) ordered on y_i for $i = 1, \cdots, n$. Zhou (1992) proposed M-estimator of the regression parameter with a quadratic error loss function.We consider the similar weighting scheme as Zhou (1992) replacing the optimal problem of LS-SVR by

$$min \frac{1}{2}\boldsymbol{\omega}'\boldsymbol{\omega} + \frac{C}{2}\sum_{i=1}^{n} u_i e_i^2 \tag{3.3}$$

over $\boldsymbol{\omega}, b, \boldsymbol{e}$ subject to $y_i - \boldsymbol{\omega}'\phi(\boldsymbol{x}_i) - b = e_i,\ i = 1, \cdots, n.$

The estimates of $\boldsymbol{\alpha}$ and b can be obtained from the linear equation:

$$\begin{pmatrix} K_s + diag(\boldsymbol{u}_s)^{-1}/C & \mathbf{1} \\ \mathbf{1}' & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha} \\ b \end{pmatrix} = \begin{pmatrix} \boldsymbol{y}_s \\ 0 \end{pmatrix} \tag{3.4}$$

where $\boldsymbol{x}_s = \{\boldsymbol{x}_i|\delta_i = 1\}$, $\boldsymbol{y}_s = \{y_i|\delta_i = 1\}$, $\boldsymbol{u}_s = \{u_i|u_i \ne 0\} = \{u_i|\delta_i = 1\}$, $K_s = K(\boldsymbol{x}_s, \boldsymbol{x}_s)$ is an $n_s \times n_s$ matrix and n_s is a size of $\boldsymbol{u}_s$. The estimated regression function of t given $\boldsymbol{x}$ is obtained as follows,

$$\widehat{m}(\boldsymbol{x}) = K(\boldsymbol{x}, \boldsymbol{x}_s)\boldsymbol{\alpha} + b. \tag{3.5}$$

Some u_i's could be inflated values, which could result in inaccurate estimation of the survival times.

The functional structures of LS-SVR with the censored data is characterized by hyper-parameters, the regularization parameter C and the kernel parameters. To select the hyper-parameters, we define the cross validation (CV) function as follows,

$$\begin{aligned} CV(\lambda) &= \frac{1}{n}\sum_{i=1}^{n} u_i(y_i - \widehat{m}(\boldsymbol{x}_i)^{(-i)})^2 = \frac{1}{n}\sum_{i\in \mathrm{Is}} u_i(y_i - \widehat{m}(\boldsymbol{x}_i)^{(-i)})^2 \\ &= \frac{1}{n}\sum_{j=1}^{n_s} u_{sj}(y_{sj} - \widehat{m}(\boldsymbol{x}_{sj})^{(-j)})^2 \end{aligned} \tag{3.6}$$

where $I_s = \{i = 1, \cdots, n|\delta_i = 1\}$, u_{sj} is the jth element of $\boldsymbol{u}_s = \{u_i, i \in I_s\}$, $\boldsymbol{\lambda}$ is the set of hyper-parameters and $\widehat{m}(\boldsymbol{x}_{sj})^{(-j)}$ is the regression function estimated without an observation corresponding to u_{sj}. Since for each candidates of parameters, $\widehat{m}(\boldsymbol{x}_{sj})^{(-j)}$ should be evaluated, selecting parameters using CV function is computationally formidable. By leaving-out-one lemma (Kimeldorf and Wahba, 1981),

$$\begin{aligned} &(y_{sj} - \widehat{m}(\boldsymbol{x}_{sj})^{(-j)}) - (y_{sj} - \widehat{m}(\boldsymbol{x}_{sj})) \\ &= \widehat{m}(\boldsymbol{x}_{sj}) - \widehat{m}(\boldsymbol{x}_{sj})^{(-j)} \approx \frac{\partial \widehat{m}(\boldsymbol{x}_{sj})}{\partial y_{sj}}(y_{sj} - \widehat{m}(\boldsymbol{x}_{sj})^{(-j)}) \end{aligned}$$

we have

$$(y_{sj} - m(\boldsymbol{x}_{sj})^{(-j)}) \approx \frac{y_{sj} - m(\boldsymbol{x}_{sj})}{1 - \dfrac{\partial m(\boldsymbol{x}_{sj})}{\partial y_{sj}}} \text{ and } m(\boldsymbol{x}_{sj}) = S_{j\cdot}\boldsymbol{y}_s,$$

where $S_{j\cdot}$ is the jth row of the hat matrix such that $S(\boldsymbol{x}_s, \boldsymbol{x}_s) = (\mathbf{1}_{ns}S_{12} + K_sS_{22})$, S_{12} and S_{22} are submatrices of the inverse matrix in (3.4),

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} = \begin{pmatrix} K_s + diag(\boldsymbol{u_s})^{-1}/C & \mathbf{1} \\ \mathbf{1}' & 0 \end{pmatrix}^{-1}. \tag{3.7}$$

Using (3.7) the ordinary cross validation (OCV) function can be obtained as follows,

$$OCV(\boldsymbol{\lambda}) = \frac{1}{n}\sum_{j=1}^{n_s} u_{sj} \left(\frac{y_{sj} - \widehat{m}(\boldsymbol{x}_{sj}|\boldsymbol{\lambda})}{1 - \dfrac{\partial \widehat{m}(\boldsymbol{x}_{sj})}{\partial y_{sj}}} \right)^2 = \frac{1}{n}\sum_{j=1}^{n_s} u_{sj} \left(\frac{y_{sj} - \widehat{m}(\boldsymbol{x}_{sj}|\boldsymbol{\lambda})}{1 - s_{jj}(\boldsymbol{\lambda})} \right)^2, \tag{3.8}$$

where s_{jj} is the jth diagonal element of the hat matrix S. By summing the weighted squared residuals in (3.8) revised by $(1 - tr(S)/n_s)^2$, GCV function can be then obtained as follows,

$$GCV(\boldsymbol{\lambda}) = \frac{1}{n}\frac{\sum_{j=1}^{n_s} u_{sj}(y_{sj} - \widehat{m}(\boldsymbol{x}_{sj}|\boldsymbol{\lambda}))^2}{(1 - tr(S)/n_s)^2}. \tag{3.9}$$

For the censored data we can obtain the estimated regression function of t given $\boldsymbol{x}$ is obtained as in (3.4), but for better estimation we propose two-step LS-SVR for in censored regression. Two-step LS-SVR for in censored regression consists of two steps as follows;

(i) Use $\{y_i, \boldsymbol{x}_i | i \in I_s\}$ as training data set and $\{\boldsymbol{x}_i | i \notin I_s\}$ as test data set to obtain the estimated regression function given $\{ \boldsymbol{x}_i | i \notin I_s\}$, $\widetilde{m}(\boldsymbol{x}_i | i \notin I_s)$.

(ii) Define pseudo response $\widetilde{y}_i$ such that $\widetilde{y}_i = y_i$ if $\delta_i = 1$ and $\widetilde{y}_i = \widetilde{m}(\boldsymbol{x}_i)$ if $\delta_i = 0$. We use $\{\widetilde{y}_i, \boldsymbol{x}_i, \delta_i, i = 1, \cdots, n\}$ as training data set and $\{\boldsymbol{x}_i, i = 1, \cdots, n\}$ as test data set to obtain the estimated regression function given $\{\boldsymbol{x}_i, i = 1, \cdots, n\}$, $\widehat{m}(\boldsymbol{x}_i)$, $i = 1, \cdots, n$. By using $\{\widetilde{y}_i, \boldsymbol{x}_i, \delta_i, i = 1, \cdots, n\}$ as training data set, we can see that $u_i = \delta_i/\widehat{G}(y_i)$, used in (3.4) is changed to $u_i^* = \delta_i/\widetilde{G}(\widetilde{y}_i)$, which affects the estimated values of $\boldsymbol{\alpha}$ and b, where $\widehat{G}$ is Kaplan-Meier estimates of the censoring distribution from $\{y_i, \delta_i, i = 1, \cdots, n\}$ and $\widetilde{G}$ is Kaplan-Meier estimates of the censoring distribution from $\{\widetilde{y}_i, \delta_i, i = 1, \cdots, n\}$. The reason why we use the pseudo response is that we empirically found that $\widetilde{y}_i$ for $i \notin I_s$ is close to the true regression function than y_i for $i \notin I_s$.

4. Numerical studies

We illustrate the performance of the censored regression estimation using two-step LS-SVR through the simulated example and real example.

Example 4.1 For the nonlinear censored regression case, 100 of x's are generated from a uniform distribution, $U(0,1)$, and (t,c)'s are generated as follows.

$$t_i = f(x_i) + \epsilon_{t_i},\ c_i = c + f(x_i) + \epsilon_{c_i},\ i = 1, \cdots, 100,$$

where $f(x_i)$ is the regression function of t_i given x_i, ϵ_{t_i}'s and ϵ_{c_i}'s are generated from normal distributions, $N(0, 0.1)$. c is chosen for 25%, 35% and 45% censoring proportion. We repeated the above procedure 100 times to obtain 100 data sets for each of various censoring proportions. The Gaussian kernel is utilized in this example, which is

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

The regularization parameter C and the kernel parameter σ^2 are obtained by GCV function (3.10). The Figure 4.1 shows the true regression function(solid line) superimposed on the scatter plots of 100 data with 25%, 35% and 45% censoring proportion, where uncensored data points are denoted by "·" and the censored by "o".

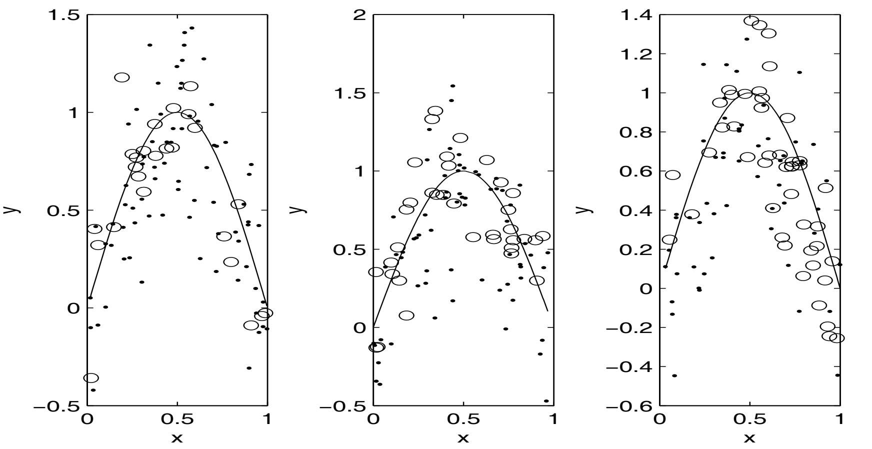

Figure 4.1 The true regression function (solid line) superimposed on the scatter plots of 100 data with various censoring proportions

Mean squared error (MSE) is used for the performance metric,

$$mse_k = \frac{1}{n}\sum_{i=1}^{n}(\widehat{f}_k(x_i) - f(x_i))^2,$$

where $f(x)$ is the true regression function, $\hat{f}_1(x)$ is the estimated regression function obtained by LS-SVRc and $\hat{f}_2(x)$ is the estimated regression function obtained by two-step LS-SVR. From 100 data sets of each censoring proportion we obtained the means, standard deviations and medians of mse_1's and mse_2's as in Table 4.1. Box plots of 100 mse_1's and mse_2's according to various censoring proportions are shown in Figure 4.2. In Table 4.2 we obtained

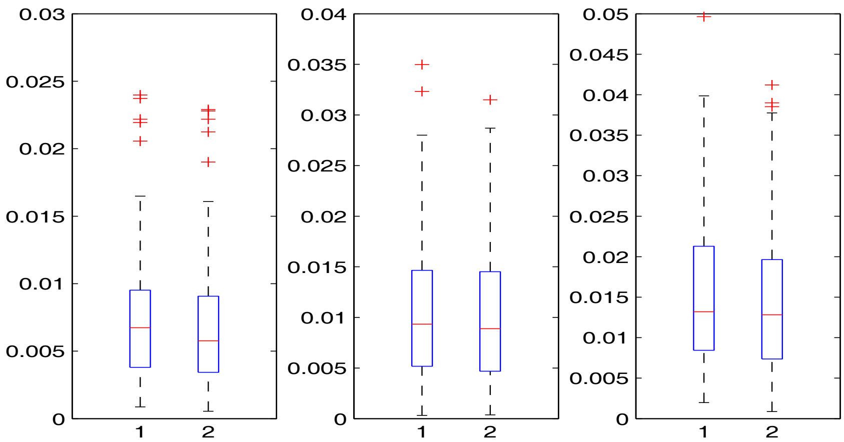

Figure 4.2 Box plots of 100 $mse_1's$ and $mse_2's$ according to various censoring proportions

Table 4.1 The means, standard deviations and medians of mse_1's and mse_2's according to various censoring proportions

censoring proportion		mse_1	mse_2
	mean	0.0074	0.0072
25%	std	0.0048	0.0048
	median	0.0066	0.0062
	mean	0.0109	0.0101
35%	std	0.0094	0.0091
	median	0.0067	0.0065
	mean	0.0158	0.0144
45%	std	0.0096	0.0090
	median	0.0131	0.0128

Table 4.2 95% confidence interval of E(mse_1)-E(mse_2) according to various censoring proportions

	25%	35%	45%
lower	0.0000376	0.0004	0.0008
upper	0.0005124	0.0010	0.0019

95% confidence interval of E(mse_1)-E(mse_2) according to various censoring proportions, we can see that all three confidence intervals includes positive numbers, which indicates that mse_2 is smaller than mse_1.

Example 4.2 We use the small cell lung cancer data in Ying *et al.* (1995). In this study, 121 patients are randomly assigned to be treated with either arm A (Cisplatin followed by Etoposide) or arm B (Etoposide followed by Cisplatin) with 62 patients assigned to arm A and 59 to arm B. Ying *et al.* (1995) considered a log-linear model in the age at entry, which is represented as

$$t_i = b + \beta_1 x_i + \eta(z_i) + \epsilon_i = b + \beta_1 x_i + \beta_2 z_i + \epsilon_i, i = 1, \cdots, 121,$$

where t_i is the base 10 logarithm of the ith patient's survival time, $x_i = 0$ if the ith patient is treated with arm A and 1 otherwise, z_i is the entry age of the ith patient. We want to evaluate their linear model using the proposed two-step LS-SVR with linear kernel. The optimal Lagrange multipliers $\{\alpha_i : i \in I_s\}$ can be obtained from the linear equation (3.4). The estimated regression function given (x_i, z_i) is obtained as

$$\widehat{m}(x_i, z_i) = \widehat{b} + \sum_{j \in I_s} \alpha_j K((x_j, z_j), (x_i, z_i)) = \widehat{b} + \widehat{\beta}_1 x_i + \widehat{\beta}_2 z_i, i = 1, \cdots, 121.$$

If K is the linear kernel $\widehat{\beta}_1 = \sum_{j \in I_s} x_j \alpha_j$ and $\widehat{\beta}_2 = \sum_{j \in I_s} z_j \alpha_j$. The estimated regression functions superimposed on the scatter plots of data points are shown in Figure 4.3. For two-step LS-SVR with linear kernel, the parameter $C = 1$ is chosen by GCV function (3.10) and estimates of (b, β_1, β_2) are obtained as (2.9960, -0.1721, -0.0038). Ying *et al.* (1995) obtained them as (3.028, -0.163, -0.004). Thus, estimates of the covariate effects from two methods are reasonably close. With the jackknife method 95% confidence intervals of b, β_1 and β_2 are obtained as (2.6552, 3.3362), (-0.3381, -0.0062), and (-0.0102, 0.0026), respectively. These results indicate that there is a significant evidence that arm A is better than arm B. For the two-step LS-SVR with Gaussian kernel, the parameters (C, σ^2) are chosen as (1, 5) by GCV function. The estimated regression functions by two-step LS-SVR are described in Figure 4.3, in there we see the linear kernel looks appropriate for these data.

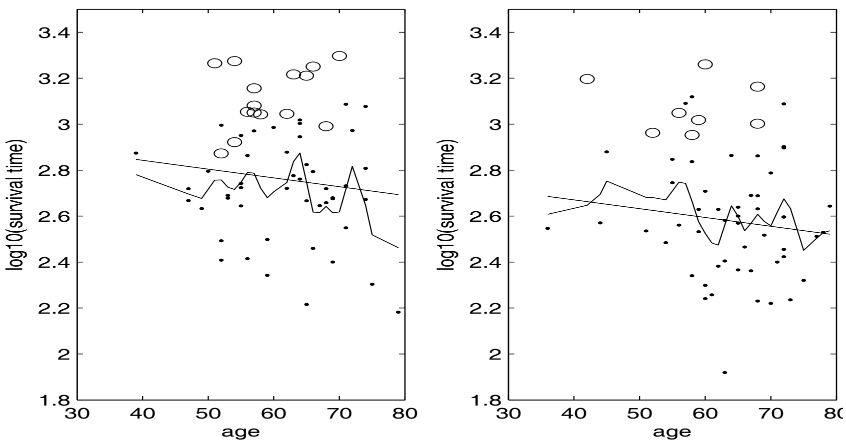

Figure 4.3 The estimated regression functions (Linear: linear kernel, Nonlinear: Gaussian kernel) for arm A (Left) and arm B (Right) by two-step LS-SVR

5. Concluding remarks

In this paper, we dealt with estimating the regression function using two-step LS-SVR when the responses are subject to randomly right censoring. We obtained GCV function for the proposed procedure. By using GCV function the model selection becomes easier and faster than that by a leave-one-out cross validation. Through the examples we showed that the proposed procedure derives the satisfying results and is attractive approaches to modelling of the censored data.

References

Cho, D. H., Shim, J. and Seok, K. H. (2010). Doubly penalized kernel method for heteroscedastic autoregressive data. *Journal of the Korean Data & Information Science Society*, **21**, 155-162.

Gunn, S. R. (1998). *Support vector machines for classification and regression*, Technical report, Department of Electronics and Computer Science, Southamption, England, United Kingdom.

Heuchenne, C. and Van Keilegom, I. (2005). Nonlinear regression with censored data, *Technometrics*, **49**, 34-44.

Hwang, C. and Shim, J. (2010). Semiparametric support vector machine for accelerated failure time model. *Journal of the Korean Data & Information Science Society*, **21**, 467-477.

Jin, Z., Lin, D. Y., Wei, L. J. and Ying, Z. L. (2003). Rank-based inference for the accelerated failure time model. *Biometrika*, **90**, 341-353.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**, 457-481.

Kimeldorf, G. and Wahba, G. (1981). Some results on Tchebycheffian spline functions. *Journal of Mathematical Analysis and Applications*, **33**, 82-95.

Koul, H., Susarla, V. and Van Ryzin J. (1981). Regression analysis with randomly right censored data. *The Annal of Statistics*, **9**, 1276-1288.

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.

Miller, R. G. (1976). Least squares regression with censored data. *Biometrika*, **63**, 449-464.

Shim, J., Kim, C. R. and Hwang, C. (2011). Semiparametric least squares support vector machine for accelerated failure time model. *Journal of the Korean Statistical Society*, **40**, 75-83.

Shim, J. (2005). Censored kernel ridge regression. *Journal of the Korean Data & Information Science Society*, **16**, 1045-1052.

Shim, J. and Lee, J. T. (2009). Kernel method for autoregressive data. *Journal of the Korean Data & Information Science Society*, **20**, 467-4720.

Smola, A. and Scholkopf, B. (1998). On a kernel-based method for pattern recognition, regression, approximation and operator inversion. *Algorithmica*, **22**, 211-231.

Stute, W. (1993). Consistent estimation under random censorship when covariables are available. *Journal of Multivariate Analysis*, **45**, 89=103.

Suykens, J. A. K. and Vanderwalle, J. (1999). Least square support vector machine classifier. *Neural Processing Letters*, **9**, 293-300.

Vapnik, V. N. (1998). *Statistical learning theory*, Wiley, New York.

Yang, S. (1999). Censored median regression using weighted empirical survival and hazard functions. *Journal of the American Statistical Association*, **94**, 137-145.

Ying, Z. L., Jung, S. H. and Wei, L. J. (1995). Survival analysis with median regression models, *Journal of the American Statistical Association*, **90**, 178-184.

Zhou, M. (1992). M-estimation in censored linear models. *Biometrika*, **79**, 837-841.