Journal of the Korean Data & Information Science Society
2010, **21**(4), 765–775

한국데이터정보과학회지

Semiparametric support vector machine for accelerated failure time model†

Changha Hwang[1] · Jooyong Shim[2]

[1]Department of Statistics, Dankook University
[2]Department of Applied Statistics, Catholic University of Daegu

Received 7 May 2010, revised 30 June 2010, accepted 5 July 2010

Abstract

For the accelerated failure time (AFT) model a lot of effort has been devoted to develop effective estimation methods. AFT model assumes a linear relationship between the logarithm of event time and covariates. In this paper we propose a semiparametric support vector machine to consider situations where the functional form of the effect of one or more covariates is unknown. The proposed estimating equation can be computed by a quadratic programming and a linear equation. We study the effect of several covariates on a censored response variable with an unknown probability distribution. We also provide a generalized approximate cross-validation method for choosing the hyper-parameters which affect the performance of the proposed approach. The proposed method is evaluated through simulations using the artificial example.

Keywords: Accelerated failure time, generalized approximate cross validation function, hyper-parameters, semiparametric regression model, support vector machine.

1. Introduction

Among popular models in analyzing failure time data are the Cox proportional hazards (PH) model and the accelerated failure time (AFT) model. The AFT model is an appealing alternative to the widely-used PH model, in which the logarithm, or a monotonic transformation of the survival time is modelled linearly in the covariates (Cox, 1972; Kalbfleisch and Prentice, 2002). In general, the estimation of this model is carried out assuming a parametric error distribution. However, some authors have developed semiparametric estimation methods of AFT models with an unspecified error distribution to avoid the need for this parametric assumption. Two semiparametric methods have been mainly used in practice. One method is the Buckley-James estimator which adjusts censored observations using the

† 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 (2009-0086922).

1 Department of Statistics, Dankook University, 126 Jukjeon-dong, Suji-gu, Yongin, Gyeonggi 448-701, Korea.
2 Corresponding author: Department of Applied Statistics, Catholic University of Daegu, Gyungbuk 702-701, Korea. E-mail: ds1631@hamail.net

Kaplan-Meier estimator in the least squares regression. The other is the linear-rank-test-based estimator which is motivated from the score function of the partial likelihood. See for comprehensive expositions and references, Buckley and James (1979), Jones (1997), Tsiatis (1990), Wei *et al.* (1990), Jin *et al.* (2003), Chen *et al.* (2005) and Shim (2005a).

Support vector machine (SVM) has been very successful in classification and regression problems (Vapnik, 1998). See for references, Gunn (1998), Shim (2005b), Kim *et al.* (2008), Shim and Seok (2008). In this paper we propose the semiparametric SVM for the AFT model using a weighting system formed with the Kaplan-Meir estimator of censoring distribution. The use of the Kaplan-Meier weights to account for censoring has been first proposed by Stute (1993). The existing semiparametric 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 medium and high dimensional covariates.

In this paper we propose two versions of semiparametric SVMs for estimating AFT model - a semiparametric SVM for censored data and a semiparametric SVM for censored data using the iteratively reweighted least squares (IRWLS) procedure. The rest of this paper is organized as follows. In Section 2 we give a brief overview of 0-insensitive SVM for median regression. In Section 3 we propose a semiparametric SVM for estimating AFT model with the Kaplan-Meier weights for censored data and a generalized approximate cross-validation (GACV) method for determining the hyper-parameters which affect the performance of the proposed model. In Section 4 we propose a semiparametric SVM for estimating AFT model using IRWLS procedure and a GACV method. In Section 5 we present simulation studies to illustrate our methods. Finally, we present the conclusions in Section 6.

2. Support vector machine for median regression

Let the training data set denoted by $(\boldsymbol{x}_i, y_i)_{i=1}^n$, with each input $\boldsymbol{x}_i \in R^d$ and the response $y_i \in R$, where the output variable y_i is related to the input vector $\boldsymbol{x}_i$. Here the feature mapping function $\phi(\cdot) : R^d \to R^{d_f}$ maps the input space to the higher dimensional feature space where the dimension d_f is defined in an implicit way. An inner product in feature space has an equivalent kernel function in input space, $\phi(\boldsymbol{x}_i)'\phi(\boldsymbol{x}_j) = K(\boldsymbol{x}_i, \boldsymbol{x}_j)$ (Mercer, 1909). We consider the nonlinear median regression case, in which the regression function of the response given $\boldsymbol{x}$, $m(\boldsymbol{x})$, can be regarded as a nonlinear regression function of input vector $\boldsymbol{x}$.

With the absolute loss function, the estimator of the median can be defined as any solution to the optimization problem,

$$\min \frac{1}{2}\boldsymbol{w}'\boldsymbol{w} + C\sum_{i=1}^{n} |y_i - m(\boldsymbol{x}_i)|. \tag{2.1}$$

We can express the regression problem by formulation for SVM as follows.

$$\min \frac{1}{2}\boldsymbol{w}'\boldsymbol{w} + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{2.2}$$

subject to

$$y_i - \boldsymbol{w}'\phi(\boldsymbol{x}_i) - b \le \xi_i \qquad (2.3)$$
$$\boldsymbol{w}'\phi(\boldsymbol{x}_i) + b - y_i \le \xi_i^*,\ \xi_i, \xi_i^* \ge 0$$

where $\boldsymbol{w}$ is a weight vector, $\phi(\boldsymbol{x}_i)$ is a feature mapping function, b is a bias, and C is a regularization parameter penalizing the training errors. We construct a Lagrange function as follows:

$$L = \frac{1}{2}\boldsymbol{w}'\boldsymbol{w} + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) - \sum_{i=1}^{n}\alpha_i(\xi_i - y_i + \boldsymbol{w}'\phi(\boldsymbol{x}_i) + b) \qquad (2.4)$$
$$- \sum_{i=1}^{n}\alpha_i^*(\xi_i^* + y_i - \boldsymbol{w}'\phi(\boldsymbol{x}_i) - b) - \sum_{i=1}^{n}(\eta_i\xi_i + \eta_i^*\xi_i^*).$$

We notice that the positivity constraints $\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \ge 0$ should be satisfied. After taking partial derivatives of equation (2.4) with regard to the primal variables $(\boldsymbol{w}, \xi_i, \xi_i^*)$ and plugging them into equation (2.4), we have the optimization problem below.

$$\max -\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(\boldsymbol{x}_i, \boldsymbol{x}_j) + \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)y_i$$

with constraints

$$\sum_{i}^{n}(\alpha_i - \alpha_i^*) = 0 \text{ and } \alpha_i, \alpha_i^* \in [0, C],$$

where the data points corresponding to positive values of α_i or α_i^* are called support vectors. Solving the above equation with the constraints determines the optimal Lagrange multipliers, α_i, α_i^*, the estimator of the median given the input vector $\boldsymbol{x}$ are obtained as follows.

$$\widehat{m}(\boldsymbol{x}) = \widehat{b} + K(\boldsymbol{x}, \boldsymbol{x})(\hat{\boldsymbol{\alpha}} - \widehat{\boldsymbol{\alpha}}^*), \qquad (2.5)$$

where $\widehat{b}$ is obtained via Kuhn-Tucker conditions (Kuhn and Tucker, 1951) such as,

$$\widehat{b} = \frac{1}{n_s}\sum_{i \in I_s}(y_i - K(\boldsymbol{x}_i, \boldsymbol{x})(\hat{\boldsymbol{\alpha}} - \widehat{\boldsymbol{\alpha}}^*)), \qquad (2.6)$$

with n_s a size of the set $I_s = \{i = 1, \cdots, n | 0 < \hat{\alpha}_i < C, 0 < \hat{\alpha}_i^* < C\}$.

In the nonlinear case, $\boldsymbol{w}$ is no longer explicitly given. However, it is uniquely defined in the weak sense by the dot products. Here the linear regression model can be regarded as the special case of the nonlinear regression model by using identity feature mapping function, that is, $\phi(\boldsymbol{x}) = \boldsymbol{x}$ which implies the linear kernel such that $K(\boldsymbol{x}_1, \boldsymbol{x}_2) = \boldsymbol{x}_1'\boldsymbol{x}_2$.

3. Semiparametric SVM with censored data

In this section, without loss of generality, we assume the input vector included in the parametric part of the regression function is known to have the linear effect on the response variable.

Let t_i be the response variables corresponding to vector, $\boldsymbol{x}_i$ or transformation on it, where $i = 1, 2, \cdots, n$. Let $(\boldsymbol{x}_i, \boldsymbol{z}_i)$ be the associated vector of covariates with (p, q) components. Let $m(\boldsymbol{x}_i, \boldsymbol{z}_i)$ be the regression function of the response variable given $(\boldsymbol{x}_i, \boldsymbol{z}_i)$.

We assume that $m(\boldsymbol{x}_i, \boldsymbol{z}_i)$ is related to the vector of covariates $(\boldsymbol{x}_i, \boldsymbol{z}_i)$ in a semiparametric form as

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

where $\boldsymbol{\beta}$ is a $(p \times 1)$ regression parameter vector and $\phi(\boldsymbol{z}_i)$ 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 $\boldsymbol{x}_i$ for $i = 1, 2, \cdots, n$. c_i's are assumed to be independently distributed with unknown survival distribution functions G.

In most practical cases 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 $m(\boldsymbol{x}_i, \boldsymbol{z}_i)$ based on $(\delta_1, y_1, \boldsymbol{x}_1, , \boldsymbol{z}_1), \cdots, (\delta_n, y_n, \boldsymbol{x}_n, \boldsymbol{z}_n)$. 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 Meier, 1958) of survival distribution function G of c_i's can be obtained as,

$$\begin{cases} \widehat{G}(y) = \prod_{i:y_{(i)} \leq y} \left(\dfrac{n-i}{n-i+1} \right)^{1-\delta_{(i)}} & y \leq y_{(n)} \\ 0 & \text{otherwise} \end{cases} \tag{3.3}$$

where $(y_{(i)}, \delta_{(i)})$ is (y_i, δ_i) ordered on y_i for $i = 1, 2, \cdots, n$. Zhou (1992) proposed M-estimator of the regression parameter with a quadratic loss function.

We consider the similar weighting scheme as Zhou (1992) replacing the optimal problem of SVM (2.1) by

$$\min \frac{1}{2}\boldsymbol{w}'\boldsymbol{w} + C\sum_{i=1}^{n} u_i(\xi_i + \xi_i^*) \tag{3.4}$$

over $\{\boldsymbol{w}, b, \boldsymbol{\beta}\}$ subject to

$$\begin{aligned} y_i - b - \boldsymbol{\beta}'\boldsymbol{x}_i - \boldsymbol{w}'\phi(\boldsymbol{z}_i) &\leq \xi_i, \\ b + \boldsymbol{\beta}'\boldsymbol{x}_i + \boldsymbol{w}'\phi(\boldsymbol{z}_i) - y_i &\leq \xi_i^*, \end{aligned}$$

where $u_i = \delta_i/(\hat{G}(y_i))$ is the weight proposed by Koul *et al.* (1981) imposed on the i th observation. Thus, we can construct Lagrangian function as follows:

$$L = \frac{1}{2}\boldsymbol{w}'\boldsymbol{w} + C\sum_{i=1}^{n} u_i(\xi_i + \xi_i^*) - \sum_{i=1}^{n} \alpha_i(\xi_i - y_i + b + \boldsymbol{\beta}'\boldsymbol{x}_i + \boldsymbol{w}'\phi(\boldsymbol{z}_i)) \\ - \sum_{i=1}^{n} \alpha_i^*(\xi_i^* + y_i - b - \boldsymbol{\beta}'\boldsymbol{x}_i - \boldsymbol{w}'\phi(\boldsymbol{z}_i)) - \sum_{i=1}^{n} (\eta_i\xi_i + \eta_i^*\xi_i^*). \tag{3.5}$$

Here, (α_i, α_i^*)'s are Lagrange multipliers. Taking partial derivatives of Lagrangian function (3.5) with regard to the primal variables $(\boldsymbol{w}, b, \boldsymbol{\beta}, \boldsymbol{\xi}^{(*)})$, we have

$$\boldsymbol{w} = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\phi(\boldsymbol{z}_i)', \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0, \\ \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\boldsymbol{x}_i' = \boldsymbol{0}, \alpha_i, \alpha_i^* \in [0, u_iC]. \tag{3.6}$$

From the equation (3.6), we have $\alpha_i = 0$ and $\alpha_i^* = 0$ if $u_i = 0$. We rewrite Lagrangian function (3.5) as follows:

$$L = \frac{1}{2}\boldsymbol{w}'\boldsymbol{w} + C\sum_{i\in \mathrm{I_s}} u_i(\xi_i + \xi_i^*) - \sum_{i\in \mathrm{I_s}} \alpha_i(e + \xi_i - y_i + b + \boldsymbol{\beta}'\boldsymbol{x}_i + \boldsymbol{w}'\phi(\boldsymbol{z}_i)) \\ - \sum_{i\in \mathrm{I_s}} \alpha_i^*(e + \xi_i^* + y_i - b - \boldsymbol{\beta}'\boldsymbol{x}_i - \boldsymbol{w}'\phi(\boldsymbol{z}_i)) - \sum_{i\in \mathrm{I_s}} (\eta_i\xi_i + \eta_i^*\xi_i^*).$$

where $\mathrm{I}_s = \{i = 1, \cdots, n | u_i \neq 0\}$.

Thus we have the optimization problem with uncensored data as follows:

$$\max -\frac{1}{2}\sum_{i,j\in \mathrm{I_s}} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(\boldsymbol{z}_i, \boldsymbol{z}_j) + \sum_{i\in \mathrm{I_s}} (\alpha_i - \alpha_i^*)y_i \tag{3.7}$$

with constraints

$$\sum_{i\in \mathrm{Is}}(\alpha_i - \alpha_i^*) = 0, \sum_{i\in \mathrm{Is}}(\alpha_i - \alpha_i^*)\boldsymbol{x}_i' = \boldsymbol{0}, \alpha_i, \alpha_i^* \in [0, u_iC]. \tag{3.8}$$

The estimator of the regression function given the input vector $\boldsymbol{x}$ and $\boldsymbol{z}$ are obtained as follows:

$$\widehat{m}(\boldsymbol{x}, \boldsymbol{z}) = \hat{b} + \hat{\boldsymbol{\beta}}'\boldsymbol{x} + K(\boldsymbol{z}, \boldsymbol{z})(\widehat{\boldsymbol{\alpha}} - \widehat{\boldsymbol{\alpha}}^*), \tag{3.9}$$

where $K(\boldsymbol{z}, \boldsymbol{z})$ is the kernel function constructed with $\boldsymbol{z}_i$ for $i = 1, \cdots, n$.

Here $\widehat{b}$ and $\widehat{\boldsymbol{\beta}}$ are obtained via Kuhn-Tucker conditions (Kuhn and Tucker, 1951) such as,

$$\begin{pmatrix} \widehat{b} \\ \widehat{\boldsymbol{\beta}} \end{pmatrix} = \left(X_{sv}{}'X_{sv}\right)^{-1} X_{sv}{}' \left(\boldsymbol{y}_{sv} - K(\boldsymbol{z}_{sv}, \boldsymbol{z})(\widehat{\boldsymbol{\alpha}} - \widehat{\boldsymbol{\alpha}}^*)\right), \tag{3.10}$$

where $I_{sv} = \{i = 1, \cdots, n | u_i \neq 0, 0 < \hat{\alpha}_i < u_i C, 0 < \hat{\alpha}_i^* < u_i C\}$, $\boldsymbol{y}_{sv}$ consists of y_i's with $i \in I_{sv}$, $\boldsymbol{z}_{sv}$ consists of z_i's with $i \in I_s$, and X_{sv} consists of $(1, \boldsymbol{x}_i)$'s with $i \in I_{sv}$. We can see that $I_{sv} = \{i = 1, \cdots, n | y_i - \widehat{\mu}(\boldsymbol{x}_i, \boldsymbol{z}_i) = 0\}$.

The functional structures of the semiparametric SVM 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:

$$CV(\theta) = \frac{1}{n}\sum_{i=1}^{n} u_i |y_i - \hat{m}(\boldsymbol{x}_i, \boldsymbol{z}_i)^{(-i)}|, \tag{3.11}$$

where θ is the set of parameters and $\widehat{m}(\boldsymbol{x}_i, \boldsymbol{z}_i)^{(-i)}$ is the regression function estimated without i th observation. Since for each candidates of parameters, $\widehat{m}(\boldsymbol{x}_i, \boldsymbol{z}_i)^{(-i)}$ for $i = 1, \cdots, n$, should be evaluated, selecting parameters using CV function is computationally formidable. If we assume that we can express $\hat{m}(\boldsymbol{x}, \boldsymbol{z})$ as the linear product of the hat matrix and $\boldsymbol{y}$, GACV function can be written as follows by Yuan (2006):

$$GACV(\lambda) = \frac{\sum_{i=1}^{n} u_i |y_i - \widehat{m}(\boldsymbol{x}_i, \boldsymbol{z}_i)|}{n - trace(H)}, \tag{3.12}$$

where $\hat{m}(\boldsymbol{x}, \boldsymbol{z}) = H\boldsymbol{y}$ with the (i, j)th element $h_{ij} = \partial \hat{m}(\boldsymbol{x}_i, \boldsymbol{z}_i)/\partial y_j$.

From Li *et al.* (2007) we have that the trace of the hat matrix H equals to the sum of size of set I_{sv} used in (3.10) since $h_{ii} = \partial \hat{m}(\boldsymbol{x}_i, \boldsymbol{z}_i)/\partial y_i = 1$ for $i \in I_{sv}$. Thus we have GACV function as follows;

$$GACV(\lambda) = \frac{1}{(n - n_{sv})}\sum_{i=1}^{n} u_i |y_i - \widehat{m}(\boldsymbol{x}_i, \boldsymbol{z}_i)|, \tag{3.13}$$

where n_{sv} is the size of I_{sv}.

4. Semiparametric SVM with censored data using IRWLS procedure

In this section we consider a differentiable objective function, then we can have faster computing and easy derivation of generalized approximate cross validation (GACV) function. To have the objective function differentiable, we use the modified 0-insensitive loss function $\rho_\delta(\cdot)$ which is attained by providing the differentiability at 0 by differing from the original 0-insensitive loss function in the small interval $(-\delta, \delta)$ for sufficiently small $\delta > 0$,

$$\rho_\delta(r) = -r\mathrm{I}(r \leq -\delta) + \frac{1}{4}r^2\mathrm{I}(-\delta < r \leq \delta) + r\mathrm{I}(r > \delta). \tag{4.1}$$

Now the problem (3.4) becomes obtaining $(b, \boldsymbol{\beta}, \boldsymbol{\alpha})$ to minimize

$$L(b, \boldsymbol{\beta}, \boldsymbol{\alpha}) = \frac{1}{2}\boldsymbol{\alpha}' K(\boldsymbol{z}, \boldsymbol{z})\boldsymbol{\alpha} + C\sum_{i=1}^{n} u_i \rho_\delta(y_i - b - \boldsymbol{x}_i'\boldsymbol{\beta} - K(\boldsymbol{z}_i, \boldsymbol{z})\boldsymbol{\alpha}) \tag{4.2}$$

Taking partial derivatives of (4.2) with regard to $(b, \boldsymbol{\beta}, \boldsymbol{\alpha})$ leads to the optimal values of $(b, \boldsymbol{\beta}, \boldsymbol{\alpha})$ to be the solution to

$$\begin{aligned} 0 &= \mathbf{1}'UW\boldsymbol{y}_s - \mathbf{1}'UW\boldsymbol{x}\boldsymbol{\beta} + \mathbf{1}'UWK\boldsymbol{\alpha} + \mathbf{1}'UW\mathbf{1}b \\ \mathbf{0}_{p\times 1} &= \boldsymbol{x}'UW\boldsymbol{y} - \boldsymbol{x}'UW\boldsymbol{x}\boldsymbol{\beta} - \boldsymbol{x}'UWK\boldsymbol{\alpha} - \boldsymbol{x}'UW\mathbf{1}b \\ \mathbf{0}_{n\times 1} &= K\boldsymbol{\alpha} - CKUW\boldsymbol{y} + CKUW\boldsymbol{x}\boldsymbol{\beta} + CKUWK\boldsymbol{\alpha} + CKUW\mathbf{1}b. \end{aligned} \tag{4.3}$$

$K = K(\boldsymbol{z}, \boldsymbol{z})$, U is the diagonal matrix of non-zero u_i's and W is a diagonal matrix w_{ii}, $i = 1, \cdots, n$ obtained from the derivative of the modified absolute loss function as

$$w_{ii} = -\frac{1}{r_i}\mathrm{I}(r_i \le -\delta) + \frac{2}{\delta}\mathrm{I}(-\delta < r_i \le \delta) + \frac{1}{r_i}\mathrm{I}(r_i > \delta), \tag{4.4}$$

where $r_i = y_i - b - \boldsymbol{x}_i\boldsymbol{\beta} - K(\boldsymbol{z}_i, \boldsymbol{z})\boldsymbol{\alpha}$.

The solution to the equations (4.3) cannot be obtained in a single step since W contains $(b, \boldsymbol{\beta}, \boldsymbol{\alpha})$ therein. Thus we need to apply IRWLS procedure which starts with initialized values $(b^{(0)}, \boldsymbol{\beta}^{(0)}, \boldsymbol{\alpha}^{(0)})$ as follows:

1) Calculate W from (4.4) using $r_i = y_i - b^{(l)} - \boldsymbol{x}_i\boldsymbol{\beta}^{(l)} - K(\boldsymbol{z}_i, \boldsymbol{z})\boldsymbol{\alpha}^{(l)}$ 2) Obtain $(b^{(l+1)}, \boldsymbol{\beta}^{(l+1)}, \boldsymbol{\alpha}^{(l+1)})$ from

$$\begin{pmatrix} b^{(l+1)} \\ \boldsymbol{\beta}^{(l+1)} \\ \boldsymbol{\alpha}^{(l+1)} \end{pmatrix} = \begin{pmatrix} \mathbf{1}'UW\mathbf{1}, & \mathbf{1}'UW\boldsymbol{x}_s, & \mathbf{1}'UWK \\ \boldsymbol{x}_s'UW\mathbf{1}, & \boldsymbol{x}_s'UW\boldsymbol{x}_s, & \boldsymbol{x}_s'UWK \\ KUW, & KUW\boldsymbol{x}_s, & KUWK + K/C \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{1}'UW \\ \boldsymbol{x}_s'UW \\ KUW \end{pmatrix} \boldsymbol{y}_s. \tag{4.5}$$

3) Iterate steps until convergence.

The estimator of the regression function given the input vector $\boldsymbol{x}$ and $\boldsymbol{z}$ are obtained as follows:

$$\widehat{m}(\boldsymbol{x}, \boldsymbol{z}) = \hat{b} + \hat{\boldsymbol{\beta}}'\boldsymbol{x} + K(\boldsymbol{z}, \boldsymbol{z})\widehat{\boldsymbol{\alpha}}. \tag{4.6}$$

To select the hyper-parameters, we define the cross validation (CV) function as follows:

$$CV(\theta) = \frac{1}{n}\sum_{i=1}^{n} u_i \rho_\delta(y_i - m(\boldsymbol{x}_i, \boldsymbol{z}_i)^{(-i)}), \tag{4.7}$$

where θ is the set of hyper-parameters and $m(\boldsymbol{x}_i, \boldsymbol{z}_i)^{(-i)}$ is the regression function estimated without an observation corresponding to u_i. Since for each candidates of parameters, $m(\boldsymbol{x}_i, \boldsymbol{z}_i)^{(-i)}$ should be evaluated, selecting parameters using CV function is computationally formidable. By leaving-out-one lemma (Kimeldorf and Wahba, 1971),

$$\begin{aligned} &(y_i - m(\boldsymbol{x}_i, \boldsymbol{z}_i)^{(-i)}) - (y_i - m(\boldsymbol{x}_i, \boldsymbol{z}_i)) \\ &= m(\boldsymbol{x}_i) - m(\boldsymbol{x}_i, \boldsymbol{z}_i)^{(-i)} \approx \frac{\partial m(\boldsymbol{x}_i, \boldsymbol{z}_i)}{\partial y_i}(y_i - m(\boldsymbol{x}_i, \boldsymbol{z}_i)^{(-i)}) \end{aligned}$$

we have

$$(y_i - m(\boldsymbol{x}_i, \boldsymbol{z}_i)^{(-i)}) \approx \frac{y_i - m(\boldsymbol{x}_i, \boldsymbol{z}_i)}{1 - \dfrac{\partial m(\boldsymbol{x}_i, \boldsymbol{z}_i)}{\partial y_i}} \text{ and } m(\boldsymbol{x}_i, \boldsymbol{z}_i) = S_{i.}\boldsymbol{y}, \tag{4.8}$$

where $S_{i.}$ is the i th row of the hat matrix such that $S(\boldsymbol{x}, \boldsymbol{x}) = (\mathbf{1}_n S_{13} + \boldsymbol{x} S_{23} + K S_{33})$, S_{13} $=(1 \times n)$ matrix, S_{23} $=(p \times n)$ matrix and $S_{33} = (n \times n)$ matrix, are submatrices of the product of matrices in the right-hand side of (4.5),

$$\begin{pmatrix} S_{13} \\ S_{23} \\ S_{33} \end{pmatrix} = \begin{pmatrix} \mathbf{1}'UW\mathbf{1}, & \mathbf{1}'UW\boldsymbol{x}, & \mathbf{1}'UWK \\ \boldsymbol{x}'UW\mathbf{1}, & \boldsymbol{x}'UW\boldsymbol{x}, & \boldsymbol{x}'UWK \\ KUW, & KUW\boldsymbol{x}, & KUWK + K/C \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{1}'UW \\ \boldsymbol{x}'UW \\ KUW \end{pmatrix}. \tag{4.9}$$

Using (4.8) the approximate cross validation (ACV) function can be obtained as follows,

$$ACV(\theta) = \frac{1}{n}\sum_{i=1}^{n} u_i \rho_\delta \left(\frac{y_i - m(\boldsymbol{x}_i, \boldsymbol{z}_i|\theta)}{1 - \dfrac{\partial m(\boldsymbol{x}_i, \boldsymbol{z}_i)}{\partial y_i}} \right) = \frac{1}{n}\sum_{i=1}^{n} u_i \rho_\delta \left(\frac{y_i - m(\boldsymbol{x}_i, \boldsymbol{z}_i|\theta)}{1 - s_{ii}} \right), \tag{4.10}$$

where s_{ii} is the i th diagonal element of the hat matrix S. By summing the weighted residuals in (4.10) revised by $(1 - tr(S)/n)$, GACV function can be then obtained as follows,

$$GACV(\theta) = \frac{1}{n} \frac{\sum_{i=1}^{n} u_i \rho_\delta (y_i - m(\boldsymbol{x}_i, \boldsymbol{z}_i|\theta))}{(1 - tr(S)/n)}. \tag{4.11}$$

5. Numerical studies

In this section we perform simulation studies to investigate the finite sample behavior of semiparametric SVM regression for AFT model with a parametric component plus a nonparametric one. We compare the performances of proposed methods - SVM1 (censored semiparametric SVM) and SVM2 (censored semiparametric SVM using IRWLS procedure) with Orbe *et al.* (2003) which uses the cubic spline method (Green and Silverman, 1994) for the estimation of nonlinear part of the regression function.

We consider a semiparametric regression model with covariate vector (x, z) under 20% or 40% censoring. For $i = 1, \cdots, 100$, x_i's are equally spaced ranging from 0 to 1, and z_i's are generated from a uniform distribution $U(0, 1)$. Also, t_i's are generated from a normal distribution with mean $b + \beta x_i + \sin(\pi z_i)$ and variance 0.1. Censoring times c_i's are generated from a uniform distribution $U(0, a_1)$, where the period a_1 is chosen for 40% censoring proportion. With a_1 fixed at this value, the follow-up period a_2 is chosen for 20% censoring proportion by $U(a_2, a_1 + a_2)$. (b, β) is set to $(1, 1)$. When patients are accrued with a Poisson distribution, a_1 denotes the accrual period, and a_2 denotes the additional follow-up period after the completion of patient accrual. In each example, 200 data sets of $\{y_i, x_i, z_i, \delta_i\}$, $i = 1, \cdots, 100$ are randomly generated. The true regression functions of y vs. x and y vs. z superimposed on the scatter plot of data points from one of 200 data sets with 20% censoring are shown in Figures 1. The true regression functions are $1 + \beta x + 2/\pi$ and

$1+0.5\beta+\sin(\pi z)$, respectively, when the nonlinear and linear parts are fixed at their means, respectively. Thus, Figure 1 describes the true regression functions $1+2/\pi+x$ (Left) and $1.5+\sin(\pi z)$ (Right). Uncensored and censored data points are denoted by dots and circles, respectively. In Figure 1 we can see that both covariate values x and z are correlated with the regression function since the regression parameter β was set at 1 in the example.

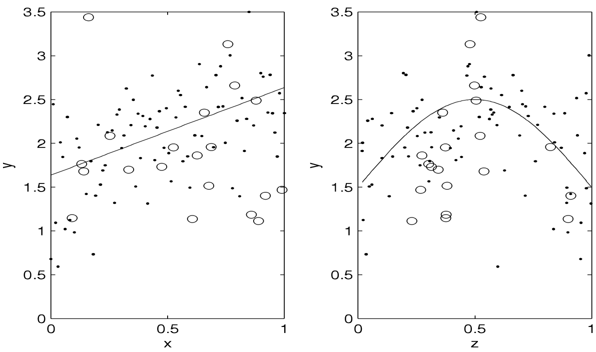

Figure 5.1 True regression functions (solid line) superimposed on the scatter plots of y vs. x (left) and y vs. z (right) from one of 200 simulated data sets for the example (β =1) with n=100 and 20% censoring.

We consider the semiparametric regression model:

$$t_i = b + x_i\beta + \eta(z_i) + \epsilon_i, i = 1, \cdots, 100,$$

where (b, β) are regression parameters to be estimated and η is an unknown nonlinear function.

The optimal Lagrange multipliers and regression estimators $(\widehat{b}, \widehat{\beta})$ can be obtained from (3.7) to (3.10) for SVM1 and (4.5) for SVM2.

The estimated regression function given (x_i, z_i) by SVM1 and SVM2 is obtained respectively, as

$$\hat{m}(x_i, z_i) = \widehat{b} + x_i\widehat{\beta} + K(z_i, \boldsymbol{z})(\widehat{\boldsymbol{\alpha}} - \widehat{\boldsymbol{\alpha}}^*),$$

and

$$\hat{m}(x_i, z_i) = \widehat{b} + x_i\widehat{\beta} + K(z_i, \boldsymbol{z})\widehat{\boldsymbol{\alpha}},$$

where $\boldsymbol{z} = (z_1, \cdots, z_{100})'$.

The Gaussian kernel is utilized in this example, which is,

$$K(z_1, z_2) = \exp\left(-\frac{(z_1 - z_2)^2}{\sigma^2}\right).$$

The regularization parameter C and the kernel parameter σ^2 are obtained by GACV function (3.13) for SVM1 and (4.11) for SVM2. Mean squared error (MSE) is used for the performance metric of the example,

$$MSE = \sum_{i=1}^{n}(b + \beta x_i + \sin(\pi z_i) - \hat{m}(x_i, z_i))^2,$$

where $\widehat{m}(x_i, z_i)$ is the estimated regression function for $i = 1, \cdots, 100$. The averages of MSEs for 20% censoring were obtained as 0.008, 0.012 and 0.049, respectively, by SVM1, SVM2 and Orbe *et al.* (2003). For 40% censoring, 0.119, 0.009 and 0.169 were obtained, respectively. We can see that proposed methods provide more accurate results than Orbe *et al.* (2003) on this example.

6. Conclusions

This paper proposes semiparametric SVM methods for the AFT model to regress a survival variable on some covariates. The proposed methods use the Kaplan-Meier weights in the objective function to account for censoring. Although we have not shown here, the proposal can be used without heavy computations even under high-dimensional covariate settings or with huge data set, since it takes over all advantages of SVM. An important issue for SVM is model selection. To this end, we provide the GACV method for choosing the hyper-parameters which affect the performance of the proposed approach. The simulation studies indicate that the proposed methods provide accurate estimates for the parametric and nonparametric components under various settings. The main advantages of the proposed approach over many existing estimation methods for the AFT model are that regression estimators can be easily computed by softwares solving a quadratic programming or a linear equation and that the proposed approach can be applied to the high dimensional covariate case. These make it easier to apply the proposed approaches to the analysis of censored data in practice.

References

Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika*, **66**, 429-436.

Chen, K., Shen, J. and Ying, Z. (2005). Rank estimation in partial linear model with censored data. *Statistica Sinica*, **15**, 767-779.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society*, **34**, 187-220.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric regression and generalized linear models*, Chapman and Hall, London.

Gunn, S. (1998). *Support vector machines for classification and regression*, ISIS Technical Report, University of Southampton.

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.

Jones, M. P. (1997). A class of semiparametric regressions for the accelerated failure time model. *Biometrika*, **84**, 73-84.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The statistical analysis of failure time data (2nd edn)* , Wiley, Hoboken.

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

Kim, M. S., Park, H. J., Hwang, C. and Shim, J. (2008). Claims reserving via kernel machine. *Journal of Korean Data & Information Science Society*, **19**, 1419-1427.

Kimeldorf, G. S. and Wahba, G. (1971). Some results on tchebycheffian spline functions. *Journal of Mathematical Analysis and its Applications*, **33**, 82-95.

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

Kuhn, H. W. and Tucker, A. W. (1951). *Nonlinear programming*, Proceedings of 2nd Berkeley Symposium, Berkeley, University of California Press, Berkeley.

Li, Y., Liu, Y. and Ji, Z. (2007). Quantile regression in reproducing kernel Hilbert spaces. *Journal of the American Statistical Association*, **102**, 255-268.

Mercer, J. (1909). Functions of positive and negative type and their connection with theory of integral equations. *Philosophical Transactions of the Royal Society*, **A**, 415-446.

Orbe, J., Ferreira, E. and Nunez-Anton, V. (2003). Censored partial regression. *Biostatistics*, **4**, 109-121.

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

Shim, J. (2005b). E-SVR using IRWLS Procedure. *Journal of Korean Data & Information Science Society*, **16**, 1087-1094.

Shim, J. and Seok, K. H. (2008). Kernel Poisson regression for longitudinal data. *Journal of Korean Data & Information Science Society*, **19**, 1353-1360.

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

Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Annals of Statistics*, **18**, 303-328.

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

Wei, L. J., Ying, Z. L. and Lin, D. Y. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika*, **77**, 845-851.

Yuan, M. (2006). GACV for quantile smoothing splines. *Computational Statistics and Data Analysis*, **50**, 813-829.

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