*Journal of the Korean Data & Information Science Society*
*2013,* **24**(1), 201–209

http://dx.doi.org/10.7465/jkdi.2013.24.1.201
한국데이터정보과학회지

# Variable selection in censored kernel regression†

Kook-Lyeol Choi[1] · Jooyong Shim[2]

[12]Department of Data Science, Inje University
Received 11 December 2012, revised 23 December 2012, accepted 2 January 2013

**Abstract**

For censored regression, it is often the case that some input variables are not important, while some input variables are more important than others. We propose a novel algorithm for selecting such important input variables for censored kernel regression, which is based on the penalized regression with the weighted quadratic loss function for the censored data, where the weight is computed from the empirical survival function of the censoring variable. We employ the weighted version of ANOVA decomposition kernels to choose optimal subset of important input variables. Experimental results are then presented which indicate the performance of the proposed variable selection method.

*Keywords*: ANOVA decomposition kernel, censored data, generalized cross validation function, kernel function, variable selection.

## 1. Introduction

Suykens and Vanderwalle (1999) proposed LS-SVM, which is a least squares version of support vector machine (SVM) originally introduced by Vapnik (1995, 1998). The solution is given by a linear system instead of a quadratic programming problem. The fact that LS-SVM has explicit primal-dual formulations has lots of advantages. Kernel tricks are used in SVM and LS-SVM to treat the nonlinear relation between input variables and output variable. See Cho *et al.* (2010), Hwang (2010a, 2010b), Shim and Lee (2009) for the reference.

The censored regression model 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 of censored regression model based on the weights proposed by Koul *et al.* (1981). Orbe *et al.* (2003) proposed the estimation procedure of censored regression model where estimators of regression parameters and nonlinear function are obtained by minimizing the penalized weighted least squares objective function through iterative method. They also proposed the procedure

† 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 (2012000646).
[1] Professor, Department of Data Science, Institute of Statistical Information, Inje University, Kimhae 621-749, Korea.
[2] Corresponding author: Adjunct professor, Department of Data Science, Institute of Statistical Information, Inje University, Kimhae 621-749, Korea. E-mail: ds1631@hanmail.net

to generate the bootstrap resamples to obtain the uncertainty measures of estimators. Jin *et al.* (2003) proposed the estimation procedure where regression parameter estimates of censored regression model are obtained from non-monotone estimating equations based on the weighted log-rank statistics. The estimating equations are solved through iterative method with Gehan (1965)-type estimate as the initial value. Ghosh and Ghosal (2006) proposed the estimation procedure based on a nonparametric Bayesian approach which uses a Dirichlet prior for the mixture of Weibull distribution in the censored regression model, where Markov Chain Monte Carlo method (Brooks, 1998) is used to obtain the marginal posterior distribution of regression parameters. Shim *et al.* (2011) proposed a semiparametric LS-SVM for the censored data using weights which Koul *et al.* (1981) proposed.

In linear regression models the stepwise regression is popular in which the selection of input variables is carried out by an automatic procedure (Draper and Smith, 1981). Many variable selection techniques for linear regression models have been extended to the context of survival models, including the best-subset selection, stepwise selection, and Bootstrap procedures (Sauerbrei and Schumacher, 1992). Recently the Lasso (least absolute shrinkage and selection operator; Tibshirani, 1997) has been proposed for Cox proportional hazards model (Cox, 1972). By shrinking some regression parameters to zero, this method provides the selection of important variables and the estimation of regression parameters simultaneously. Huang *et al.* (2005) proposed the regularization and variable selection approach using Lasso (Tibshirani, 1996). Hu and Rao (2010) proposed a weighted least squares method with censoring constraints and sparse penalization to fit censored regression models with high-dimensional covariates. There are lots of literatures in studies of variable selection: Guyon *et al.* (2002), Tibshirani *et al.* (2002), Koo *et al.* (2006). Guyon *et al.* (2002) developed SVM with a recursive features elimination algorithm and Tibshirani *et al.* (2002) developed the prediction analysis of microarrays method based upon an enhancement of the simple nearest prototype classifier. Recently, Shim *et al.* (2009) proposed the marker genes selection by the supervised weighted kernel clustering and SVM.

In this paper we propose a variable selection method in censored kernel regression, which uses the weighted ANOVA decompostion kernel. From the quadratic programming problem we obtain weights whose magnitudes imply the importance of variables on regression. The rest of paper is organized as follows. In Section 2 we present the censored kernel regression and model selection methods. In Section 3 we propose the variable selection method using the weighted ANOVA decompostion kernel. In Section 4 we perform the numerical studies with the simulated nonlinear dataset and the partially linear real dataset. In Section 5 we give the concluding remarks.

## 2. Censored kernel regression

In this paper we set $\boldsymbol{x}_i \in R^d$ be the input vector and $t_i$ be the response variable (survival time) corresponding to input vector, $\boldsymbol{x}_i$ or transformation on it, where $i = 1, \cdots, n$. In fact we cannot 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 $i = 1, \cdots, n$. $c_i$'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 the estimation of $m(\boldsymbol{x}_i)$, the regression function of the response variable given $\boldsymbol{x}_i$, based on $(\delta_1, y_1, \boldsymbol{x}_1), \cdots, (\delta_n, y_n, \boldsymbol{x}_n)$. Buckley and James (1979) defined the pseudo-response variable

$$y_i^* = y_i\delta_i + E(t_i|t_i > y_i, \boldsymbol{x}_i)(1-\delta_i). \tag{2.1}$$

They showed $E(y_i^*|\boldsymbol{x}_i) = E(t_i|\boldsymbol{x}_i)$ and proposed the iteration method to estimate the regression parameters. Koul *et al.* (1981) defined new observable responses $y_i^*$ as $y_i^* = u_i y_i$ with

$$u_i = \frac{\delta_i}{G(y_i)}, \tag{2.2}$$

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,

$$\widehat{G}(y) = \begin{cases} \prod_{i:y_{(i)} \le y} (\dfrac{n-i}{n-i+1})^{1-\delta_{(i)}} & \text{if } y \le y_{(n)} \\ 0 & \text{otherwise} \end{cases} \tag{2.3}$$

where $(y_{(i)}, \delta_{(i)})$ is $(y_i, \delta_i)$ ordered on $y_i$ for $i = 1, \cdots, n$. Koul *et al.* (1981) proposed the ordinary least squares regression of $y_i^*$ on $\boldsymbol{x}_i$. Zhou (1992) proposed the weighted least squares regression of $y_i^*$ on $\boldsymbol{x}_i$ with $u_i$ so that $(\widehat{\boldsymbol{\beta}}, \widehat{b})$ in (2.4) can be seen as the minimizer of the objective function as follows:

$$\frac{1}{2}\sum_{i=1}^{n} u_i(y_i - \boldsymbol{x}_i'\boldsymbol{\beta} - b)^2. \tag{2.4}$$

We consider the nonlinear regression case, in which the regression function of $\boldsymbol{x}_i$, $m(\boldsymbol{x}_i)$, can be regarded as a nonlinear function of input variables,

$$m(\boldsymbol{x}) = \boldsymbol{w}'\phi(\boldsymbol{x}) + b.$$

Here $b$ is a bias term and $\boldsymbol{w} \in R^{d_f}$ is a weight vector corresponding to the feature mapping function $\phi(\cdot) : R^d \to R^{d_f}$ which 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, $\phi(\boldsymbol{x}_k)'\phi(\boldsymbol{x}_l) = K(\boldsymbol{x}_k, \boldsymbol{x}_l)$ (Mercer, 1909). Using kernel tricks in SVM and LS-SVM, a penalized least squares optimization problem inspired by (2.4) can be considered with a penalty parameter $\gamma > 0$ as follows:

$$\min \frac{1}{2}\boldsymbol{\alpha}'K\boldsymbol{\alpha} + \frac{\gamma}{2}(\boldsymbol{y} - K\boldsymbol{\alpha} - b)'U(\boldsymbol{y} - K\boldsymbol{\alpha} - b), \tag{2.5}$$

where $K = K(\boldsymbol{x}, \boldsymbol{x})$, $U$ is a diagonal matrix of $u_i$'s and $\boldsymbol{\alpha}$ is an $n \times 1$ vector such that $m(\boldsymbol{x}) = K\boldsymbol{\alpha} + b$. The estimates of $(\boldsymbol{\alpha}, b)$ can be obtained from the linear equations which are results of differentiation of (2.5) with respect to $(\boldsymbol{\alpha}, b)$,

$$\begin{pmatrix} KUK + \mathrm{I}/\gamma & KU\mathbf{1} \\ \mathbf{1}'UK & \mathbf{1}'U\mathbf{1} \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha} \\ b \end{pmatrix} = \begin{pmatrix} KU \\ \mathbf{1}'U \end{pmatrix} \boldsymbol{y}. \tag{2.6}$$

Solving the linear equation (2.6), the estimated regression function given $\boldsymbol{x}_t$ is obtained as

$$\widehat{m}(\boldsymbol{x}_t) = K(\boldsymbol{x}_t, \boldsymbol{x})\widehat{\boldsymbol{\alpha}} + \widehat{b}, \tag{2.7}$$

which can be seen as the linear combination of $\boldsymbol{y}$ as follows:

$$\widehat{m}(\boldsymbol{x}_t) = (K(\boldsymbol{x}_t, \boldsymbol{x}), \mathbf{1})\begin{pmatrix} KUK + \mathrm{I}/\gamma & KU\mathbf{1} \\ \mathbf{1}'UK & \mathbf{1}'U\mathbf{1} \end{pmatrix}^{-1}\begin{pmatrix} KU \\ \mathbf{1}'U \end{pmatrix}\boldsymbol{y}. \tag{2.8}$$

For the partially linear case where the input variables can be divided into two groups - a group, $\boldsymbol{x}_1$, of variables which are related to the regression function linearly and the other group, $\boldsymbol{x}_2$, of variables which are related to the regression function nonlinearly, we can still use (2.8) by setting $K = \boldsymbol{x}_1\boldsymbol{x}_1' + K(\boldsymbol{x}_2, \boldsymbol{x}_2)$. Then the estimated regression function given $(\boldsymbol{x}_{1t}, \boldsymbol{x}_{2t})$ is obtained as

$$\widehat{m}(\boldsymbol{x}_{1t}, \boldsymbol{x}_{2t}) = (\boldsymbol{x}_{1t}\boldsymbol{x}_1 + K(\boldsymbol{x}_{2t}, \boldsymbol{x}_2))\widehat{\boldsymbol{\alpha}} + \widehat{b}. \tag{2.9}$$

The functional structure of censored kernel regression is characterized by hyper -parameters, the penalty parameter $\gamma$ and the kernel parameters. To select the optimal values of hyper-parameters of the censored kernel regression, we define a cross validation function as follows:

$$CV(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} u_i(y_i - \widehat{m}_i^{(-i)}(\boldsymbol{\theta}))^2, \tag{2.10}$$

where $\boldsymbol{\theta}$ is the set of hyper-parameters and $\widehat{m}_i^{(-i)}(\boldsymbol{\theta})$ is the predicted value of $m(\boldsymbol{x}_i)$ obtained from data without $i$ th observation. Since for each candidates of hyper-parameters, $\widehat{m}_i^{(-i)}(\boldsymbol{\theta})$ for $i = 1, \cdots, n$, should be evaluated, selecting parameters using CV function is computationally formidable.

By leaving-out-one lemma (Kimeldorf and Wahba, 1991) and the first order Taylor expansion, we have a ordinary cross validation (OCV) function (Shim *et al.*, 2011),

$$OCV(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} u_i\left(\frac{y_i - \widehat{m}_i(\boldsymbol{\theta})}{1 - s_{ii}/n}\right)^2, \tag{2.11}$$

where $s_{ii}$ is the $i$ th diagonal element of $S$ which is the hat matrix such that

$$\widehat{\boldsymbol{m}}(\boldsymbol{\theta}) = S\boldsymbol{y} = (K, \mathbf{1})\begin{pmatrix} KUK + \mathrm{I}/\gamma & KU\mathbf{1} \\ \mathbf{1}'UK & \mathbf{1}'U\mathbf{1} \end{pmatrix}^{-1}\begin{pmatrix} KU \\ \mathbf{1}'U \end{pmatrix}\boldsymbol{y}.$$

By replacing $(1 - s_{ii}/n)$ with $1 - trace(S)/n$ we have a generalized cross validation (GCV) function,

$$GCV(\boldsymbol{\theta}) = \frac{n\sum_{i=1}^{n} u_i(y_i - \widehat{m}_i(\boldsymbol{\theta}))^2}{(n - trace(S))^2}. \tag{2.12}$$

## 3. Variable selection

The ANOVA decomposition kernels are inspired by ANOVA in Statistics, which can be seen as the sum of kernels constructed by different subsets of variables (Vapnik, 1998). The ANOVA decomposition kernel has two main advantages (Saunders *et al.*, 1998) - (i) improving predictive performance by considering the different subsets as group together like variables (ii) avoiding overfitting training data by only considering some subsets of input variables.

We assume that $p$ is the prespecified number of input variables to be selected then the ANOVA decomposition kernel is defined as follows:

$$K_p = \sum_{k=1}^{d_p} K(\boldsymbol{x}_{:k}, \boldsymbol{x}_{:k}), \tag{3.1}$$

where $\boldsymbol{x}_{:k}$ is the $n \times p$ submatrix of $\boldsymbol{x}$ consisting of the $k$ th subset of $I_p = \{(k_1, \cdots, k_p) | 1 \leq k_1 < \cdots < k_p \leq d\}$ and $d_p = \binom{d}{p}$ is the size of $I_p$.

In this paperwe modify the ANOVA decompostion kernel into the weighted version such as

$$K_A = \sum_{k=1}^{d} v_k K(\boldsymbol{x}_{.k}, \boldsymbol{x}_{.k}), \tag{3.2}$$

where $\boldsymbol{x}_{.k}$ is the $k$ th column of $\boldsymbol{x}$, $v_k \geq 0$, $\sum_{k=1}^{d} v_k = 1$ and $v_k$ is a weight representing the influence of $k$ th input variable on the response. The important input variables can be selected according to magnitude of $v_k's$ such that the input variables corresponding to $p$ largest $v_k's$ are selected as $p$ most important input variables.

To select the prespecified number of input variables we need to find $v_k$'s first. But $v_k$'s cannot be obtained in a step but by the iterative procedure since $(\boldsymbol{\alpha}, b)$ contains $v_k$. The variable selection procedure for censored kernel regression can be carried out as follows:

(i) With $K = K(\boldsymbol{x}, \boldsymbol{x})$ find the optimal values of hyper-parameters from (2.11) and $(\boldsymbol{\alpha}, b)$ from the linear equation (2.6).

(ii) With $(\boldsymbol{\alpha}, b)$, find $\boldsymbol{v} = (v_1, \cdots, v_d)'$ from a quadratic programming problem,

$$\min \frac{1}{2} \boldsymbol{v}' A' U A \boldsymbol{v} - (\boldsymbol{y}' U A - b \mathbf{1}' U A - \frac{1}{2\gamma} \boldsymbol{\alpha}' A) \boldsymbol{v} \tag{3.3}$$

subject to

$$\mathbf{0} \leq \boldsymbol{v} < \mathbf{1} \text{ and } \mathbf{1}' \boldsymbol{v} = 1,$$

where $A = (K_A^1 \boldsymbol{\alpha}, \cdots, K_A^d \boldsymbol{\alpha})$ is the $n \times d$ matrix with $K_A^k = K(\boldsymbol{x}_{.k}, \boldsymbol{x}_{.k})$. The quadratic programming problem (3.3) is obtained by expressing the following optimal problem with respect to $v_k$'s,

$$\min \frac{1}{2} \boldsymbol{\alpha}' \mathbf{K}_A \boldsymbol{\alpha} + \frac{\gamma}{2} (\boldsymbol{y} - \mathbf{K}_A \boldsymbol{\alpha} - b \mathbf{1})' U (\boldsymbol{y} - \mathbf{K}_A \boldsymbol{\alpha} - b \mathbf{1})$$

subject to

$$\mathbf{0} \leq \boldsymbol{v} < \mathbf{1} \text{ and } \mathbf{1}'\boldsymbol{v} = 1.$$

(iii) Find $(\boldsymbol{\alpha}, b)$ from (2.6) by replacing $\boldsymbol{K}$ with $\mathbf{K}_A$ which is updated with newly obtained $v_k$'s in (ii) as (3.2).

(iv) Iterate (ii) and (iii) until $||\boldsymbol{v}^{(t)} - \boldsymbol{v}^{(t+1)}||$ converges.

## 4. Numerical studies

We illustrate the performance of the proposed method for the variable selection through the simulated data sets and the real data set. We indicate the performance of the proposed method by showing weighted sums of squared residuals (Zhou, 1998).

**Example 4.1** In this example we generate 200 data sets to indicate the performance of variable selection by showing that the proposed method agrees with the exhaustive search using the censored kernel regression on selection of two true important input variables. For each $i = 1, \cdots, 100$, $x_{i1}, \cdots, x_{i6}$ are generated from a uniform distribution, $U(0, 1)$ and $(t, c)$'s are generated as follows:

$$t_i = 1 + \exp(x_{i1} + x_{i4}) + \epsilon_{t_i},\ c_i = 1.2 + \exp(x_{i1} + x_{i4}) + \epsilon_{c_i},\ i = 1, \cdots, 100,$$

where $\epsilon_{t_i}$'s and $\epsilon_{c_i}$'s are generated from normal distribution, $N(0, 0.1^2)$. From 200 data sets the average and the standard error of the censoring percentage were obtained as 8.3% and 0.2%, respectively. We consider the nonlinear censored regression as follows:

$$m(\boldsymbol{x}_i) = \boldsymbol{w}'\phi(\boldsymbol{x}_i) + b,$$

where $\boldsymbol{x}_i = (x_{1i}, x_{2i}, \cdots, x_{6i})$. For each data set, a radial basis function kernel, $K(\boldsymbol{x}_j, \boldsymbol{x}_k) = \exp(-(\ 1/\sigma^2)||\boldsymbol{x}_j - \boldsymbol{x}_k||^2)$, is applied to the censored kernel regression and the optimal values of hyper-parameters $(\gamma, \sigma^2)$ are chosen from GCV function in (2.12). Table 4.1 shows the number of times selected by the proposed method as the most and the secondly most important input variables in 200 data sets. For the exhaustive search using the censored kernel regression, we use $\binom{6}{2}$ =15 sets of two input variables shown in the first column heading of Table 4.3. We divided each dataset into 15 sub-data sets according to 15 sets of response, $\delta_i = I(t_i \leq c_i)$ and two input variables such as $\{y_i, \delta_i, x_{i1}, x_{i2}\}_{i=1}^{100}$, $\{y_i, \delta_i, x_{i1}, x_{i3}\}_{i=1}^{100}, \cdots$, $\{y_i, \delta_i, x_{i5}, x_{i6}\}_{i=1}^{100}$, and obtained 15 weighted sums of squared residuals ($wSSR$ ; Zhou, 1998) such that for example,

$$wSSR(1,2) = \sum_{i=1}^{n} u_i(y_i - \widehat{m}(x_{i1}, x_{i2}))^2,$$

where $\widehat{m}(x_{i1}, x_{i2})$ is the regression function of $\boldsymbol{x}_i$ estimated from $\{y_i, \delta_i, x_{i1}, x_{i2}\}_{i=1}^{100}$. We computed the averages of 200 $wSSR$'s for each set of two input variables, which are shown in Table 4.3. By the proposed method we obtained averages of weights representing the influence of each input variable on the response in 200 dataset, which are shown in Table 4.2. From the tables we can see that the proposed method and the exhaustive search using the censored kernel regression agree on the variable selection of ($x_1$, $x_4$) as the two most important input variables. We also compute $wSSR$ using six input variables for each data set, whose average and standard error are obtained as 1.4223 and 0.0197, respectively, which is a little smaller than $wSSR(1,4)$ =1.7462 in Table 4.3.

**Table 4.1** Results of Example 4.1 by proposed method

| variable ($i$) | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| first | 99 | 2 | 3 | 88 | 3 | 5 |
| second | 25 | 38 | 37 | 29 | 39 | 32 |

**Table 4.2** Results of Example 4.1 by proposed method (standard error in parenthesis)

| variable ($i$) | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| weight | 0.2376<br>(0.0091) | 0.1316<br>(0.0163) | 0.1213<br>(0.0157) | 0.2409<br>(0.0095) | 0.1508<br>(0.0189) | 0.1149<br>(0.0160) |

**Table 4.3** Results of Example 4.1 by exhaustive search (standard error in parenthesis)

| variables ($i,j$) | average of $wSSR$ | variables ($i,j$) | average of $wSSR$ |
|---|---|---|---|
| (1, 2) | 75.6055 (0.8255) | (2, 6) | 147.0063 (1.6852) |
| (1, 3) | 75.3725 (0.7839) | (3, 4) | 75.0988 (0.7725) |
| (1, 4) | 1.7462 (0.0184) | (3, 5) | 147.3635 (1.7096) |
| (1, 5) | 75.3492 (0.7855) | (3, 6) | 147.0576 (1.7475) |
| (1, 6) | 75.6332 (0.8110) | (4, 5) | 75.0860 (0.7753) |
| (2, 3) | 147.2786 (1.7989) | (4, 6) | 75.6268 (0.7578) |
| (2, 4) | 75.6595 (0.7997) | (5, 6) | 146.4083 (1.6820) |
| (2, 5) | 147.2902 (1.7703) | | |

**Example 4.2** In this example we apply the proposed method to a study on multiple myeloma reported by Krall *et al.* (1975). Among a total of 65 patients who were treated with alkylating agents, 48 died during the study and 17 survived. The observed survival times were measured in months. There are nine input variables, which are described in Table 4.4. We consider the nonlinear censored regression as follows:

$$m(\boldsymbol{x}_i) = \boldsymbol{w}'\phi(\boldsymbol{x}_i) + b,$$

where $\boldsymbol{x}_i = (x_{1i}, x_{2i}, \cdots, x_{9i})$ = (Platelet, Frac, $\cdots$, SCalc). Using the radial basis kernel the regression function can be rewritten as follows:

$$m(\boldsymbol{x}_i) = K(\boldsymbol{x}_i, \boldsymbol{x})\boldsymbol{\alpha} + b.$$

$A$ in (3.3) becomes $A = (K_A^1\boldsymbol{\alpha}, K_A^2\boldsymbol{\alpha}, \cdots, K_A^9\boldsymbol{\alpha})$, where $K_A^k = K(\boldsymbol{x}_{.k}, \boldsymbol{x}_{.k})$ and $\boldsymbol{x}_{.k}$ is a $65 \times 1$ vector of the $k$ th input variable. A radial basis function kernel is applied in the censored kernel regression and the optimal values of hyper-parameters are chosen from GCV function in (2.12) as $(\gamma, \sigma^2)$ =(700, 4). By the proposed method we obtained a weight vector representing the influence of each input variable on the response as $\boldsymbol{v} = (0, 0, 0, 0.2747, 0.4721, 0, 0.0077, 0.2454, 0)'$, which implies that the most important input variables selected are rearranged in the ascending order as Age-HGB-Protein-LogPBM. For the exhaustive search using the censored kernel regression, we use $\binom{9}{k}$ sets of $k = 2, 3, 4$ input variables, $wSSR(Age, HGB)$, $wSSR(Age, HGB, Protein)$, $wSSR(Age, HGB, Protein, logPBM)$, are obtained as 0.0363, 0.005, 0.0001, respectively. We obtained $wSSR$ with all input variables as 0.0001, which is same as $wSSR(Age, HGB, Protein, logPBM)$ and implies that (Age, HGB, Protein, logPBM) are the most important input variables when we use the censored kernel regression.

**Table 4.4** Input variables of Myeloma data reported by Krall *et al.* (1975)

| variable | description | type |
|---|---|---|
| Platelet | platelets at diagnosis: 0=abnormal, 1=normal | binary |
| Frac | fractures at diagnosis: 0=none, 1=present | binary |
| logBUN | log(BUN) at diagnosis | continuous |
| HGB | hemoglobin at diagnosis | continuous |
| Age | age at diagnosis in years | continuous |
| logWBC | log(WBC) at diagnosis | continuous |
| logPBM | log(PBM) at diagnosis | continuous |
| Protein | proteinuria at diagnosis | continuous |
| SCalc | serum calcium at diagnosis | continuous |

*BUN: blood urea nitrogen WBC: white blood cells in the blood
PBM: plasma cells in bone marrow

## 5. Conclusions

In this paper, we dealt with the variable selection method in the censored kernel regression. The weights representing the influence of input variables are obtained by minimizing the quadratic programming problem composed of the weighted ANOVA decompostion kernels. The proposed method is reliable in the point that which agree with the exhaustive search in the simulated data. Through the examples we showed that the proposed method derives the good solutions to the variable selection in the censored kernel regression.

## References

Brooks, S. P. (1998). Markov Chain Monte Carlo method and its application. *The Statistician*, **47**, 69-100.

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

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.

Cox, D. R. (1972). Regression models and life tables (with discussions). *Journal of the Royal Statistical Society B*, **74**, 187-220.

Draper, N. and Smith, H. (1981). *Applied regression analysis*, 2nd Edition, John Wiley & Sons, Inc., New York.

Gehan, E. A. (1965). Generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika*, **52**, 202-223.

Ghosh, K. S. and Ghosal, S. (2006). Semiparametric accelerated failure time models for censored data. *Bayesian Statistics and its Applications*, **15**, 213-229.

Guyon, I., Weston, J., Barnhill, S. and Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, **46**, 389-422.

Hu, S. and Rao, J. S. (2010). *Sparse penalization with censoring constraints for estimating high dimensional AFT models with applications to microarray data analysis*, Technical Report 07 of Division of Biostatistics, Case Western Reserve University, Ohio.

Huang, J., Ma, S. and Xie, H. (2005). *Regularized estimation in the accelerated failure time model with high dimensional covariates,* Technical Report No. 349, Department of Statistics and Actuarial Science, The University of Iowa, Iowa.

Hwang, H. (2010a). Fixed size LS-SVM for multiclassification problems of large datasets. *Journal of the Korean Data & Information Science Society*, **21**, 561-567.

Hwang, H. (2010b). Variable selection for multiclassification by LS-SVM. *Journal of the Korean Data & Information Science Society*, **21**, 959-965.

Jin, Z., Lin, D. Y., Wei, L. J. and Ying, Z. (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 American Statistical Association*, **53**, 457-481.

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

Koo, J. Y., Sohn, I., Kim, S. and Lee, J. W. (2006). Structured polychotomous machine diagnosis of multiple cancer types using gene expression. *Bioinformatics*, **22**, 950-990.

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

Krall, J. N., Utho, V. A. and Harvey, J. B. (1975). A step-up procedure for selecting variables associated with survival. *Biometrics*, **31**, 49-57

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.

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

Sauerbrei, W. and Schumacher, M. (1992). A bootstrap resampling procedure for model building: Application to the Cox regression model. *Statistical Medicine*, **11**, 2093-2099.

Saunders, C., Gammerman, A. and Vovk, V. (1998). Ridge regression learning algorithm in dual variables. *Proceedings of the 15th International Conference on Machine Learning*, 515-521.

Shim, J. and Lee, J. T. (2009). Kernel method for autoregressive data. *Journal of the Korean Data & Information Science Society*, **20**, 467-472.

Shim, J., Kim, C. 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., Sohn, I., Kim, S., Lee, J. W., Green, P. E. and Hwang, C. (2009). Selecting marker genes for cancer classification using supervised weighted kernel clustering and the support vector machine. *Computational Statistics and Data Analysis*, **53**, 1736-1742.

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

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society B*, **58**, 267-288.

Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. *Statistics in Medicine*, **16**, 385-395.

Tibshirani, R., Hastie, T., Narasimhan, B. and Chu, G. (2002). Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proceedings of the National Academy of Sciences*, **99**, 6567-6572.

Vapnik, V. N. (1995). *The nature of statistical learning theory*, Springer, New York.

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

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

Zhou, M. (1998). *Regression with censored data : The synthetic data and least squares approach*, Technical Report 374, University of Kentucky, Kentucky.