Journal of the Korean Data & Information Science Society
2016, **27**(6), 1653–1660

http://dx.doi.org/10.7465/jkdi.2016.27.6.1653
한국데이터정보과학회지

Estimation and variable selection in censored regression model with smoothly clipped absolute deviation penalty†

Jooyong Shim[1] · Jongsig Bae[2] · Kyungha Seok[3]

[1]Department of Statistics, Institute of Statistical Information, Inje University
[2]Department of Mathematics, Sungkyunkwan University
[3]Department of Statistics, Institute of Statistical Information, Inje University

Received 28 September 2016, revised 10 November 2016, accepted 23 November 2016

Abstract

Smoothly clipped absolute deviation (SCAD) penalty is known to satisfy the desirable properties for penalty functions like as unbiasedness, sparsity and continuity. In this paper, we deal with the regression function estimation and variable selection based on SCAD penalized censored regression model. We use the local linear approximation and the iteratively reweighted least squares algorithm to solve SCAD penalized log likelihood function. The proposed method provides an efficient method for variable selection and regression function estimation. The generalized cross validation function is presented for the model selection. Applications of the proposed method are illustrated through the simulated and a real example.

Keywords: Censored regression model, generalized cross validation function, iteratively reweighted least squares procedure, smoothly clipped absolute deviation penalty, variable selection.

1. Introduction

The least squares method and censored regression model are widely accepted and well understood in the field of statistics. Koul *et al.* (1981) proposed a simple least squares method for the censored regression model by using the weighted observations. Zhou (1992) proposed an M-estimators of the censored regression model based on the Koul *et al.* (1981). Orbe *et al.* (2003) proposed an estimation procedure of the censored partial regression model. The objective function with the penalized weighted least squares through iterative procedure is used. They also proposed the bootstrap sampling to get the uncertainty measures of the estimators. Ghosh and Ghosal (2006) proposed a nonparametric Bayesian method which uses a Dirichlet prior for the mixture of Weibull distributions in the censored regression

† This research was supported by Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (NRF-2015R1D1A1A01056582). This work was supported by the 2016 Inje University research grant.

[1] Adjunct professor, Department of Statistics, Institute of Statistical Information, Inje University, Kimhae 50834, Korea.
[2] Professor, Department of Mathematics, Sungkyunkwan University, Suwon 16419, Korea.
[3] Corresponding author: Professor, Department of Statistics, Institute of Statistical Information, Inje University, Kimhae 50834, Korea. E-mail: statskh@inje.ac.kr

model. They used Markov Chain Monte Carlo method (Geyer, 1992) to obtain the marginal posterior distribution of parameters.

Generally, all the input variables may not much affect the survival time so that some regression parameters may be zeros in true regression model. Many variable selection methods such as the stepwise method, best subset method and the Bootstrap procedures (Sauerbrei and Schumacher, 1992) for the linear regression models have been extended to the survival models. Tibshirani (1997) applied LASSO (least absolute shrinkage and selection operator; Tibshirani, 1996) to the Cox proportional hazards model (Cox, 1972). It is well known from Tibshirani (1996) that LASSO performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the statistical model it produces. Huang *et al.* (2005) proposed a regularized estimation in the accelerated failure time model using LASSO. Hu and Rao (2010)proposed a weighted least squares method for the censored regression model with sparse penalization.

Smoothly clipped absolute deviation (SCAD) penalty proposed by Fan and Li (2001) is known to satisfy all of the three desirable requirements (unbiasedness, sparsity, continuity) for penalty functions. They also show that the performance of SCAD is expected to be as good as that of the oracle estimator as the sample size increases. We consider the censored regression with the SCAD penalty. We apply the iteratively reweighted least squares (IRWLS) procedure to solve the local linear approximation of SCAD penalized log-likelihood function. Then the variable selection is performed by using the absolute values of estimated parameters.

The paper is organized as follows. In Section 2 we give the reviews of censored regression. In Section 3 we present an estimation method for censored regression with SCAD penalty. In Section 4 we conduct the numerical studies with simulated and a real data set. Finally we give the conclusions in Section 5.

2. Censored regression

In this paper we set $\boldsymbol{x}_i$ be the input vector of size $d_x \times 1$ and t_i be the response variable (survival times) corresponding to input vector, $\boldsymbol{x}_i$ or transformation on the response variable, where $i = 1, 2, \cdots, n$. In fact we cannot observe t_i's but the observed variable, $y_i = \min(t_i, c_i)$ and $\delta_i = I\ (t_i \le c_i)$, where $I(\cdot)$ denotes the indicator function and c_i is the censoring variable corresponding to $\boldsymbol{x}_i$. Here c_i 's are assumed to be independently distributed with unknown survival distribution function, S . We set $f(\boldsymbol{x}_i)$ be the regression function of the response variable given $\boldsymbol{x}_i$. We assume that $f(\boldsymbol{x}_i)$ is related to the input vector $\boldsymbol{x}_i$ in a linear form without a bias as

$$f(\boldsymbol{x}_i) = \boldsymbol{x}_i'\boldsymbol{\beta}, i = 1, 2, \cdots, n,$$

where $\boldsymbol{\beta}$ is a $d_x \times 1$ regression parameter vector.

In most practical cases the survival distribution function of c_i 's, S, is not known and its estimate is usually obtained by the Kaplan-Meier (1958) estimator or its variants. The problem considered in the censored regression model is that of the estimation of $f(\boldsymbol{x}_i)$ based on $(\delta_1, y_1, \boldsymbol{x}_1), \cdots, (\delta_n, y_n, \boldsymbol{x}_n)$. Buckley and James (1979) proposed the pseudo-response variable such that

$$\widetilde{y}_i = y_i\delta_i + E(t_i|t_i > y_i, \boldsymbol{x}_i)(1 - \delta_i).$$

They showed $E(y_i^*|\boldsymbol{x}_i) = E(t_i|\boldsymbol{x}_i)$ and developed an iteration procedure to estimate the regression parameters $\boldsymbol{\beta}$. Koul *et al.* (1981) proposed the new observable response y_i^* as $y_i^* = w_i y_i$ with

$$w_i = \frac{\delta_i}{\widehat{S}(y_i)},$$

and showed $\widetilde{y}_i$ has the same mean as t_i and thus it follows the same linear model as t_i does. Here, $\widehat{S}$, the Kaplan-Meier estimators of survival distribution function S of c_i 's can be obtained as,

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

where $(y_{(i)}, \delta_{(i)})$ is (y_i, δ_i) ordered on y_i for $i = 1, \cdots, n$. Koul *et al.* (1981) proposed the least squares regression of y_i^* on $\boldsymbol{x}_i$ in the censored regression model. Zhou (1992) obtained the estimators of regression parameters using the weighted least squares regression as follows:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'W\boldsymbol{X})^{-1}\boldsymbol{X}'W\boldsymbol{y}, \tag{2.1}$$

where $\boldsymbol{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)'$ is an $n \times d_x$ matrix and W is a diagonal matrix of w_i 's. $\widehat{\boldsymbol{\beta}}$ in (2.1) can be regarded as the minimizer of the following objective function

$$\ell(\boldsymbol{\beta}) = \frac{1}{2}\sum_{i=1}^{n} w_i(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2. \tag{2.2}$$

3. SCAD penalized censored regression

From (2.2) we assume that $r_i = \sqrt{w_i}(y_i - f(\boldsymbol{x}_i))$ follows a probability distribution such that $p(r_i) \propto \exp(-0.5r_i^2)$. Then the negative loglikelihood of the given data set without constant terms can be expressed as

$$\ell(\boldsymbol{f}) = \frac{1}{2}\sum_{i=1}^{n} w_i(y_i - f(\boldsymbol{x}_i))^2,$$

where $f(\boldsymbol{x}_i) = \boldsymbol{x}_i'\boldsymbol{\beta}$. Then the maximum likelihood estimators of $\boldsymbol{\beta}$ are obtained by minimizing (2.2).

The maximum likelihood estimators of $\boldsymbol{\beta}$ generally lead severe overfitting, which inspires us to use a penalty term of $\boldsymbol{\beta}$ to avoid overfitting. Among many penalty functions, the SCAD is rated as the best one. The SCAD penalty results in small parameter estimators being set to zero, a few other parameter estimators being shrunk toward to zero while retaining the large parameter estimators as they are, which produce the sparse estimators of parameters. We use SCAD penalty in the censored regression, which is given by

$$p_\lambda(|\beta_k|) = \lambda|\beta_k|I(|\beta_k| \le \lambda) - \frac{\beta_k^2 - 2a\lambda|\beta_k| + \lambda^2}{2(a-1)}I(\lambda < |\beta_k| \le a\lambda) + \frac{(a+1)\lambda^2}{2}I(|\beta_k| > a\lambda),$$

where $\lambda > 0$ is a penalty parameter. Fan and Li (2001) suggested that $a = 3.7$ is a good choice for various problems. Then the penalized objective function of $\boldsymbol{\beta}$ is obtained as follows:

$$L_0(\boldsymbol{\beta}) = \frac{1}{2}\sum_{i=1}^{n} w_i(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2 + \sum_{k=1}^{d_x} p_\lambda(|\beta_k|), \tag{3.1}$$

where $p_\lambda(|\beta_k|)$ is a SCAD penalty.

Since the objective function $L_0(\boldsymbol{\beta})$ in (3.1) is not differentiable with respect to $\boldsymbol{\beta}$ at zero, we need a modification of $L_0(\boldsymbol{\beta})$ for the differentiability. We first use the local linear approximation of SCAD penalty as follows:

$$p_\lambda(|\beta_k|) \approx p_\lambda(|\beta_k^{(0)}|) + p_\lambda^{(1)}(|\beta_k^{(0)}|)(|\beta_k| - |\beta_k^{(0)}|) for |\beta_k| \approx |\beta_k^{(0)}|,$$

where $p_\lambda^{(1)}(|\beta_k^{(0)}|)$ is the first derivative of $p_\lambda(|\beta_k|)$ with respect to $|\beta_k|$ at $|\beta_k^{(0)}|$, which is given by

$$p_\lambda^{(1)}(|\beta_k|) = \lambda I(|\beta_k| \le \lambda) + \frac{(a\lambda - |\beta_k|)_+}{(a-1)} I(|\beta_k| > \lambda)$$

with $(r)_+ = rI(r \ge 0)$.

Then the objective function $L(\boldsymbol{\beta})$ in (3.1) can be modified to

$$L(\boldsymbol{\beta}) = \frac{1}{2}\sum_{i=1}^{n} w_i(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2 + \sum_{k=1}^{d_x} p_\lambda^{(1)}(|\beta_k|)|\beta_k|. \tag{3.2}$$

The objective function $L(\boldsymbol{\beta})$ in (3.2) is convex in $\boldsymbol{\beta}$, but still not differentiable with respect to $\boldsymbol{\beta}$ at zero. Using (3.2) we define an objective function given $\boldsymbol{\beta}^*$ as follows:

$$L(\boldsymbol{\beta}|\boldsymbol{\beta}^*) = \frac{1}{2}\sum_{i=1}^{n} w_i(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2 + \frac{1}{2}\sum_{k=1}^{d_x} p_\lambda^{(1)}(|\beta_k^*|)(\frac{\beta_k^2}{|\beta_k^*|} + |\beta_k^*|).$$

Note that $L(\boldsymbol{\beta}|\boldsymbol{\beta}^*) \ge L(\boldsymbol{\beta})$, with equality if $\boldsymbol{\beta} = \boldsymbol{\beta}^*$ (Krishnapuram *et al.*, 2005), and is differentiable with respect to $\boldsymbol{\beta}$. With the objective function $L(\boldsymbol{\beta}|\boldsymbol{\beta}^*)$ we propose a variable selection algorithm for SCAD penalized censored regression as follows:

i) Set $\widehat{\boldsymbol{\beta}}^{(0)}$ be the solution to $\ell(\boldsymbol{\beta})$ in (2.2).

ii) $\widehat{\boldsymbol{\beta}}^{(t+1)}$ is the minimizer of

$$L(\boldsymbol{\beta}|\widehat{\boldsymbol{\beta}}^{(t)}) = \frac{1}{2}\sum_{i=1}^{n} w_i(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2 + \frac{1}{2}\sum_{k=1}^{d_x} p_\lambda^{(1)}(|\widehat{\beta}_k^{(t)}|)(\frac{\beta_k^2}{|\widehat{\beta}_k^{(t)}|} + |\widehat{\beta}_k^{(t)}|),$$

with respect to $\boldsymbol{\beta}$ as

$$\widehat{\boldsymbol{\beta}}^{(t+1)} = (\boldsymbol{X}'W\boldsymbol{X} + V(\widehat{\boldsymbol{\beta}}^{(t)}))^{-1}\boldsymbol{X}'W\boldsymbol{y},$$

where $\widehat{\beta}_k^{(t)}$ is the estimator obtained at the t th iteration and $V(\widehat{\boldsymbol{\beta}}^{(t)})$ is the diagonal matrix consisted of $p_\lambda^{(1)}(|\widehat{\beta}_k^{(t)}|)/|\widehat{\beta}_k^{(t)}|, k = 1, \cdots, d_x$.

iii) Iterate ii) until convergence.

iv) Build the new data set composed of 0 and $|\widehat{\beta}_k|$'s for $k = 1, \cdots, d_x$.

v) Using $|\widehat{\beta}_k|$ and $\max\{|\widehat{\beta}_k|\} - |\widehat{\beta}_k|$, divide the new data set into two clusters.

vi) Variables corresponding to the cluster which includes $\max\{|\widehat{\beta}_k|\}$ are treated as the important variables selected.

The SCAD penalty in the censored regression is affected by the penalty parameter λ . We define the cross validation (CV) function to choose the optimal penalty parameter as follows:

$$CV(\lambda) = \frac{1}{n} \sum_{i=1}^{n} w_i (y_i - \widehat{f}_\lambda^{(-i)}(\boldsymbol{x}_i))^2.$$

Here $\widehat{f}_\lambda^{(-i)}(\boldsymbol{x}_i)$ is the estimated function without the i th observation. Since $f_\lambda^{(-i)}(\boldsymbol{x}_i)$ for $i = 1, \cdots, n$, should be calculated for each candidate of penalty parameter, choosing parameters through CV function is computationally expensive. To save computing time, we use the GCV function is obtained as follows:

$$GCV(\lambda) = \frac{n \sum_{i=1}^{n} w_i (y_i - \widehat{f}_\lambda(\boldsymbol{x}_i))^2}{(n - tr(H))^2},$$

where $H = \boldsymbol{x}(\boldsymbol{x}'W\boldsymbol{x} + V(\lambda))^{-1}\boldsymbol{x}'W$ such that $\widehat{f}_\lambda(\boldsymbol{x}) = H\boldsymbol{y}$ with the (i,j) th element $h_{ij} = \partial \widehat{f}(\boldsymbol{x}_i)/\partial y_j$. We can find the details of derivation of GCV function in Shim and Seok (2014).

4. Numerical studies

This section investigates capabilities of the proposed method (censored regression with SCAD penalty, CREG_SCAD) for function estimation and variable selection methods through the simulated as well as a real data set.

4.1. Artificial data

We generate 200 artificial data sets to compare the performance of the estimation and the variable selection with the L1-penalized censored regression (CREG_L1; Hwang *et al.*, 2011) and the weighted least squares regression (WLSE) of Zhou (1992) in (2.1). For each $i = 1, \cdots, 100$, $x_{i1}, \cdots, x_{i20}$ are independently generated from a uniform distribution, $U(0,1)$, respectively, and (t,c) 's are generated as follows:

$$t_i = f(\boldsymbol{x}_i) + \epsilon_{t_i},\ c_i = 0.4 + \epsilon_{c_i},\ i = 1, \cdots, 100,$$

where $f(\boldsymbol{x}_i) = \boldsymbol{x}_i'\boldsymbol{\beta}$, $\boldsymbol{\beta} = (1, 0, -1, 0, \cdots, 0)' \in R^{20\times 1}$, ϵ_{t_i}'s and ϵ_{c_i}'s are independently generated from normal distribution, $N(0, 0.5^2)$. For each data set, the optimal values of the penalty parameters of CREG_SCAD and CREG_L1 are chosen from GCV function.

Average of 200 censoring proportions is obtained as 0.3194. Figure 4.1 shows the box plots of each $\widehat{\beta}_k$'s by CREG_SCAD (left), CREG_L1 (middle) and WLSE of Zhou (1992) (right). From Figure 4.1 we can see that 3 models show that x_1 and x_3 are the most important variables, and we also can see that CREG_SCAD provides the stable estimates of β_k's of zero true values. Table 4.1 shows average numbers of selected variables and average numbers of selected true important variables (x_1, x_3).

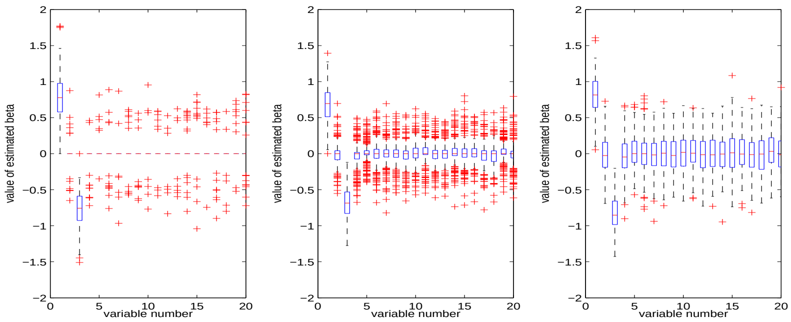

Figure 4.1 Box plots of regression parameter estimators in example 4.1

Recovery rate is defined as the ratio of average number of selected true important variables to the average number of selected variables as in Kim *et al.* (2013). From Table 4.1 we know that the recovery rate of CREG_SCAD is higher than that of CREG_L1. We also obtained the mean squared error of each data set. As results we obtained the averages of 200 mean squared errors and their standard errors for CREG_SCAD as (0.0828, 0.0034), (0.0971, 0.0024) for CREG_L1, and (0.1096, 0.0027) for WLSE of Zhou (1992), respectively. This implies that CREG_SCAD has better estimation performance than CREG_L1 and WLSE in this example. In Table 4.2 we obtained the averages of $\widehat{\beta}_k$'s for $k = 1, \cdots, 20$. The boldfaced figure in each column signifies the estimate closest to true value of β_k . From Figure 4.1, Table 4.1 and 4.2 we can see that CREG_SCAD shows better performances of variable selection than other two methods.

Table 4.1 Average number of selected variables and average number of selected true important variables (standard error in parenthesis)

	Avg. no. of selected variables	Avg. no. of selected important variables	recovery rate
CREG_SCAD	2.620(0.0319)	1.710(0.0162)	0.6527
CREG_L1	13.670(0.1723)	1.990(0.0071)	0.1456

Table 4.2 Estimations of parameters in example 4.1 (standard error in parenthesis)

parameter	true value	CREG_SCAD	CREG_L1	WLSE
β_1	1	0.7272 (0.0281)	0.6743 (0.0186)	**0.8084** (0.0189)
β_2	0	**-0.0008** (0.0086)	-0.0075 (0.0123)	-0.0169 (0.0177)
β_3	-1	-0.7150 (0.0247)	-0.6857 (0.0161)	**-0.8272** (0.0170)
β_4	0	**-0.0213** (0.0088)	-0.0292 (0.0135)	-0.0358 (0.0196)
β_5	0	0.0106 (0.0061)	**0.0100** (0.0109)	0.0196 (0.0164)
β_6	0	-0.0079 (0.0094)	**0.0009** (0.0139)	0.0022 (0.0199)
β_7	0	**-0.0073** (0.0082)	-0.0075 (0.0131)	-0.0090 (0.0177)
β_8	0	**-0.0009** (0.0091)	0.0012 (0.0131)	0.0027 (0.0189)
β_9	0	**-0.0104** (0.0078)	-0.0179 (0.0127)	-0.0201 (0.0177)
β_{10}	0	**-0.0001** (0.0084)	0.0045 (0.0128)	0.0087 (0.0176)
β_{11}	0	**0.0038** (0.0081)	0.0151 (0.0126)	0.0216 (0.0176)
β_{12}	0	**-0.0058** (0.0069)	-0.00107 (0.0126)	-0.0060 (0.0190)
β_{13}	0	-0.0043 (0.0072)	-0.0059 (0.0121)	**0.0017** (0.0181)
β_{14}	0	0.0072 (0.0099)	-0.0150 (0.0134)	**-0.0016** (0.0181)
β_{15}	0	0.0150 (0.0115)	0.0149 (0.0141)	**0.0134** (0.0200)
β_{16}	0	0.0110 (0.0071)	**0.0097** (0.0121)	0.0186 (0.0178)
β_{17}	0	**-0.0069** (0.0094)	-0.0141 (0.0132)	-0.0158 (0.0186)
β_{18}	0	**-0.0063** (0.0068)	-0.0193 (0.0123)	-0.0149 (0.0185)
β_{19}	0	**0.0174** (0.0088)	0.0291 (0.0121)	0.0502 (0.0176)
β_{20}	0	**0.0004** (0.0108)	-0.0019 (0.0136)	0.0004 (0.0188)

4.2. Real data

We applied CREG_SCAD to the Diffuse large B-cell lymphoma survival times and gene expression data set by Rosenwald *et al.* (2002). This data set consists of 7399 genes expression values from 240 patients with Diffuse large B-cell lymphoma. Deaths of 138 patients are included during the follow-ups with a median death time of 2.80 years.

As in Bair and Tibshirani (2004), the patients were randomly divided into training and test sets with 160 patients and 80 patients, respectively. Each input variable x_{ik}, $i = 1, \cdots, 240$, $k = 1, \cdots, 7399$, is standardized to $(x_{ik} - \min(\boldsymbol{X}_{.k}))/(\max(x_k) - \min(\boldsymbol{X}_{.k}))$, where $\boldsymbol{X}_{.k}$ is the k th column of the 240×7399 input matrix $\boldsymbol{X}$.

To compare the performance of variable selection and estimation method for censored data, Li (2006), Hu and Rao (2010) proposed to divide patients into high risk and low risk group according to the survival time estimators, and then compare the difference between survival times of two risk groups, where the estimated medians are used as the cut-points. We can say that the selected variables are expected to be predictive if there is a significant difference between two risk groups.

The estimated medians were obtained by CREG_SCAD as 4.0882 for training data and 3.7977 for test data. Using these cut-points we obtained Kaplan-Meier estimates of survival functions for two risk groups of training data and test data. They are shown in Figure 4.2. The dotted lines are survival functions for a high risk group, the solid lines are those for a low risk group, and 'o' represents censored data point.

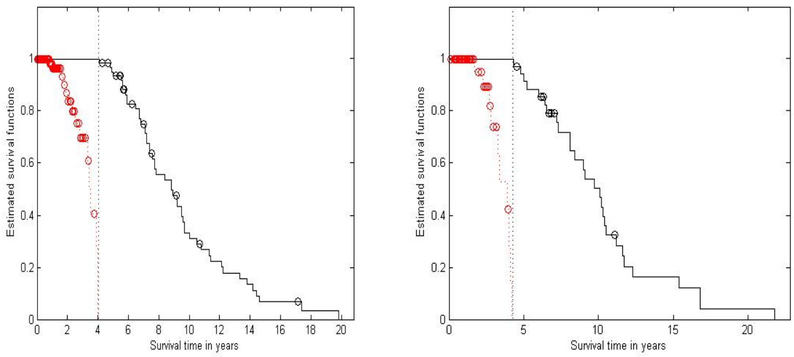

Figure 4.2 Survival function estimators for two risk groups of training data (left) and test data (right)

By quadratic programming with censoring and LASSO constraints (Hu and Rao, 2010) 79 genes were selected, log-rank tests show that $\chi^2_{(1)} = 116.12$ and p-value $< 1\text{e-}10$ for training data, and $\chi^2_{(1)} = 4.55$ and p-value $= 0.017$ for test data. Li (2006) reported that by the gradient LASSO (Huang *et al.*, 2005) 37 genes were chosen and resulted in p-value of the log-rank test of test data equal to 0.05. By CREG_L1, 99 genes were chosen, log-rank tests show that $\chi^2_{(1)} = 32.17$ and p-value $< 1\text{e-}7$ for training data, and $\chi^2_{(1)} = 12.96$ and p-value $=0.00031$ for test data. By CREG_SCAD, 5 genes were selected, log-rank tests reveal that $\chi^2_{(1)} = 56.77$ and p-value$< 1\text{e-}10$ for training data, and $\chi^2_{(1)} = 32.45$ and p-value$< 1\text{e-}7$ for test data, which show highly significant difference in survival times between the two risk groups for training data and test data. Thus the smaller p-values of log-rank tests show better predictive performance of CREG_SCAD than the other methods for test data.

5. Conclusions

In this paper, we proposed the regression function estimation and variable selection method based on SCAD penalized censored regression model. To solve the local linear approximation of SCAD penalized loglikelihood function of censored regression model we use the iteratively reweighted least squares method. It provides an efficient variable selection and the generalized cross validation function for easy model selection. Through the examples we investigated that the proposed method yields the satisfying results.

References

Bair, E. and Tibshirani, R. (2004). Semi-supervised methods to predict patient survival from gene expression data. *PLoS Biology*, **2**, 511-522.

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

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

Geyer, C. J. (1992). Practical Markov chain Monte Carlo (with discussion). *Statistical Science*, **7**, 473-511.

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

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, OH, USA.

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, IA, USA.

Hwang, C., Kim, M. and Shim, J. (2011). Variable selection in L1 penalized censored regression. *Journal of the Korean Data & Information Science Society*, **22**, 951-959.

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

Kim, J., Sohn, I., Kim, D. H., Son, D. S., Ahn, H. and Jung, S. H. (2013). Prediction of a time-to-event trait using genome wide SNP data. *BMC Bioinformatics*, **14**, 58.

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

Krishnapuram, B., Carlin, L., Figueiredo, M. A. T. and Hartermink, A. J. (2005). Sparse multinomial logistic regression: Fast algorithms and generalization bounds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27**, 957-968.

Li, H. (2006). *Censored data regression in high-dimension and low-sample size settings for genomic applications,* UPenn Biostatistics Working Paper 9, University of Pennsylvania, PA, USA.

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

Rosenwald, A., Wright, G., Chan, W. C., Connors, J. M., Campo, E., Fisher, R. I., Gascoyne, R. D., Muller-Hermelink, H. K., Smeland, E. B., Giltnane J. M. and *et al.* (2002). The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *New England Journal of Medicine*, **346**, 1937-1947.

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

Shim, J. and Seok, K. (2014). A transductive least squares support vector machine with the difference convex algorithm. *Journal of the Korean Data & Information Science Society*, **25**, 455-464.

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.

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