Genomics & Informatics Vol. 5(3) 95-101, September 2007

An Application of the Clustering Threshold Gradient Descent Regularization Method for Selecting Genes in Predicting the Survival Time of Lung Carcinomas

Seungyeoun Lee[1]* and Youngchul Kim[2]

[1]Department of Applied Mathematics, Sejong University, Seoul 143-747, Korea, [2]Department of Statistics, Seoul National University, Seoul 151-747, Korea

Abstract

In this paper, we consider the variable selection methods in the Cox model when a large number of gene expression levels are involved with survival time. Deciding which genes are associated with survival time has been a challenging problem because of the large number of genes and relatively small sample size (n << p). Several methods for variable selection have been proposed in the Cox model. Among those, we consider least absolute shrinkage and selection operator (LASSO), threshold gradient descent regularization (TGDR), and two different clustering threshold gradient descent regularization (CTGDR)—the *K*-means CTGDR and the hierarchical CTGDR—and compare these four methods in an application of lung cancer data. Comparison of the four methods shows that the two CTGDR methods yield more compact gene selection than TGDR, while LASSO selects the smallest number of genes. When these methods are evaluated by the approach of Ma and Huang (2007), none of the methods shows satisfactory performance in separating the two risk groups using the log-rank statistic based on the risk scores calculated from the selected genes. However, when the risk scores are calculated from the genes that are significant in the Cox model, the performance of the log-rank statistics shows that the two risk groups are well separated. Especially, the TGDR method has the largest log-rank statistic, and the *K*-means CTGDR method and the LASSO method show similar performance, but the hierarchical CTGDR method has the smallest log-rank statistic.

Keywords: variable selection, regularization, shrinkage estimate, LASSO, threshold gradient descent regularization, the Cox model

*Corresponding author: E-mail leesy@sejong.ac.kr
Tel +82-2-3408-3161, Fax +82-02-3408-3315
Accepted 2 September 2007

Introduction

One of main issues in survival analysis is to investigate the association of the survival time of patients with various clinical covariates. The Cox model has been most popularly used for analyzing survival data and provides predictive variables of survival time by using a variety of classical methods for variable selection, such as forward selection, backward elimination, and stepwise selection. However, classical methods such as stepwise selection procedures yield a computational problem when a large number of gene expression variables are involved in considering the association of survival time and often suffer from high variability. In addition, variable selection can be more challenging due to censoring mechanisms in survival analysis. Shrinkage methods such as LASSO have been proposed for Cox's proportional hazards model based on partial or pseudo-partial likelihoods (Tibshirani, 1996; Tibshirani, 1997). LASSO is widely used and has shown good performance. Another regularization method is the TGDR method proposed by Gui and Li (2005), which is used to identify a small number of individual genes and to build predictive models based on those genes.

On the other hand, it is well known that there exist genes whose expressions are highly correlated and should be put into clusters (Tamayo *et al.*, 1999). Cluster analysis methods have been employed in microarray studies to reduce the large number of genes into a small number of gene clusters. Once a small number of gene clusters are constructed using methods such as *K*-means or hierarchical methods, the mean expressions of genes within the same cluster are computed and then used as covariates in the final model. However, this approach has a limitation of selecting the feature at the cluster level, which implies that all genes within the selected clusters are included in the final model. Since it is not necessarily true that all genes are associated with survival time, even though genes within the same cluster may have correlated expressions, noisy genes may yield less reliable models. In order to incorporate the cluster structure into variable selection, Ma and Huang (2007) proposed a clustering threshold gradient descent regularization (CTGDR) method in which feature selection is made at both the cluster level and the individual gene level within each cluster. The CTGDR method considers the cluster

structure and takes advantage of both the cluster-based and regularized variable selection methods.

In this paper, we review four methods of variable selection and assess these methods for use in the Cox model when a large number of gene expression levels are involved in predicting survival time. Furthermore, these methods are evaluated using the log-rank statistic to compare their performance. We describe the Cox proportional hazards model and review three methods—LASSO, TGDR and CTGDR—to allow researchers to choose the most appropriate method for selecting variables. The performances of these methods are compared using a data set of lung carcinomas published on the PNAS website (www. pnas.org), and at www.genome.wi.mit.edu/ MPR/lung, and a discussion is given.

Model and Methods

The Cox proportional hazards model

Consider the survival data setup, $Y=(X, \Delta)$, where $X=\min(T, C)$ and $\Delta = I(T \leq C)$. Here, T and C denote the survival time and censoring time, respectively, and Z denotes the covariate vector. The Cox model (Cox, 1972) assumes that the conditional hazard function is independent of the time, t, given as:

$$\lambda(t \mid Z) = \lambda_0(t)\exp(\beta' Z)$$

where $\lambda_0(t)$ is the unknown baseline hazard function and β is the regression coefficient. One usually estimates the parameter β in the Cox model without specification of $\lambda_0(t)$ through maximization of the partial likelihood function, defined by:

$$L(\beta) = \prod_{i=1}^{n} \left\{ \frac{\exp(\beta' Z_i)}{\sum_{j \in r_i} \exp(\beta' Z_j)} \right\}^{\delta_i}$$

where $r_i = \{j : T_j \geq T_i\}$ is the risk set at time T_i and $\delta_i = I(T_i \leq C_i)$ is an uncensored indicator function.

Methods

We reviewed three methods for selecting susceptible genes among a large number of genes for relatively small sizes. There is a wide variety of regularization methods depending on how they define the objective functions with regularizing parameters. Here, LASSO, TGDR, and CTGDR will be compared in the study of the association between survival time and gene expression levels in the Cox model.

LASSO

Denote the log partial likelihood function by $\ell(\beta) = \log L(\beta)$ and assume that the covariates Z_{ij} are standardized so that $\sum_i Z_{ij}/n = 0$ and. $\sum_i Z_{ij}^2/n = 1$. In the linear regression setting, Tibshirani (1996) proposed minimization of the residual sum of squares, subject to a constraint of the form $\sum|\beta_j| \leq s$ and called the resulting procedure LASSO .Therefore, the estimator of β is obtained via the following criterion:

$$\beta = \arg\min \ell(\beta) \text{ subject to } \sum|\beta_j| \leq s$$

where $s > 0$ is a user-specified parameter. Suppose that β_{j^0} are the estimates that maximize the partial likelihood. Then, if $\sum|\beta_j^0| \leq s$, the solutions are the usual partial likelihood estimates. If $\sum|\beta_j| \geq s$, however, the solutions are shrunken toward zero. The attractive feature of LASSO is that quite often some of the solution coefficients are exactly zero, whereas the ridge regression approach shrinks coefficients but does not give coefficients that are exactly zero. LASSO is a tool for achieving a parsimonious model and provides a more interpretable final model.

As described in Tibshirani (1997), the estimation procedure of LASSO is expressed by the usual Newton-Raphson update as an iterative reweighted least squares step, replacing the weighted least squares step by a constrained weighted least squares procedure. Let H denote the design matrix of regressor variables and $\eta = \mathrm{H}\beta$, define $u = \frac{\partial \ell}{\partial \eta}$, $A = -\frac{\partial^2 \ell}{\partial \eta \eta^T}$ and $y = \eta + A^{-1}u$ (see Hastie and Tibshirani, 1990). Using a one-term Taylor series expansion for $\ell(\beta) = (y-\eta)^T A(y-\eta)$, we solve the problem. Then,the solution of LASSO is obtained by iterating the following procedure given by Tibshirani (1997):

1. Fix s and initialize $\hat{\beta} = 0$.
2. Compute η, u, A and y based on the current value of $\hat{\beta}$.
3. Minimize $(y-\mathrm{H}\beta)^T A(y-\mathrm{H}\beta)$ subject to $\sum|\beta_i| \leq s$.
4. Repeat step 2 and 3 until $\hat{\beta}$ does not change.

One difficulty with the above procedure is that *A* is a full of matrix, and hence it requires computation of $O(N^2)$ elements. To avoid this, we replace *A* with a diagonal matrix D that has the same diagonal elements as *A*. As argued by Hastie and Tibshirani (1990, pp.212-213), the diagonal elements of *A* are larger than the off-diagonal elements, and hence the modified algorithm should behave similarly to the original one.

TGDR

In fitting data based on linear models, the gradient descent pathfinding paradigm can be generalized to include the use of a wide variety of loss criteria, leading to robust methods for regression and classification, as well as to apply user-defined constraints on the parameter values. The paths induced by ridge regression (RR), gradient regularization (GD), and LASSO differ in how they define the interior points along their respective paths. With gradient descent-based procedures, one way to direct the path toward parameter points with more diverse component values is to increase the diversity of the factor values with a threshold parameter. The following procedure explains briefly how to perform threshold gradient descent regularization.

Denote $\triangle v$ as the small positive increment as in ordinary gradient descent methods (see Friedman and Popescu, 2004). In the implementation of this algorithm, we choose $\triangle \nu = 1 \times 10^{-4}$. Denote $\nu_k = k \times \triangle \nu$ as the index for the point along the parameter path after k steps. Let $\beta(\nu_k)$ denote the parameter estimate corresponding to ν_k. For any fixed threshold, $0 \le \tau \le 1$, the TGDR algorithm consists of the following steps:

1. Initialize β(0)=0 and $\nu_0 = 0$.
2. With current estimate β, compute the negative gradient $g(\nu) = -\partial L(\beta)/\partial\beta$. Denote the j-th component of $g(\nu)$ as $g_j(\nu)$. If $\max_j\{|g_j(\nu)|\}$, stop the iterations.
3. Compute the threshold vector $f(\nu)$ of length d, where the j-th component of $f(\nu) : f_j(\nu) = I\{|g_j(\nu)| \ge \tau \times \max_l |g_l(\nu)\}$
4. Update $\beta(\nu + \triangle\nu) = \beta(\nu) - \triangle\nu \times g(\nu) \times f(\nu)$ and update ν by $\nu + \triangle\nu$, where the product of f and g is component-wise.
5. Steps 2-4 are repeated κ times. The number of iterations κ is determined by cross validation.

Here, the property of β is determined by tuning parameters τ and κ. For example, β is dense even for small values of κ for $\tau \approx 0$, while β is sparse for small values of κ and remains so for a relatively large number of iterations, but will become dense eventually for $\tau \approx 1$. For the extreme case of τ = 1, the TGDR method usually increases in the direction of a single covariate in each iteration. For the middle range of τ, the characteristics of β are between those for τ=0 and τ=1. For $\tau \ne 0$, variable selection can be achieved with cross-validated, finite κ, by having certain components of β exactly zero.

CTGDR

While the TGDR method deals with individual gene selection but does not take into account the cluster structure, the CTGDR method considers both individual gene selection and cluster structure. Ma and Huang (2007) discussed two naïve CTGDR algorithms. The first naïve CTGDR method modifies step 3 of TGDR as follows:

$$f_j^1(\nu) = I\left\{\sum_{m \in C(j)} |g_m(\nu)| \ge \tau_1 \times \max_{C(k)} \sum_{l \in C(k)} |g_l(\nu)|\right\},$$

where $0 \le \tau_1 \le 1$ is the threshold tuning parameter. This algorithm uses cluster gradients to replace the individual gradients and considers the combined effects of genes in the same clusters, which implies that the genes within the same clusters may have different contributions in the final model, while all genes within the same clusters have equal contributions to the final model in traditional cluster-based methods.

The second naïve CTGDR method modifies step 3 by replacing f in step 3 of TGDR with

$$f_j^2(\nu) = I\left\{|g_j(\nu)| \ge \tau_2 \times \max_{l \in C(j)} |g_l(\nu)|\right\},$$

so that each gene is compared only with other genes within the same cluster and only important genes from each cluster are selected. This algorithm is roughly equivalent to carrying out the TGDR method in each cluster separately, and the final model includes genes selected from all clusters.

In summary, the first naïve CTGDR method carries out cluster selection but does not select important genes with each cluster, while the second naïve CTGDR method carries out gene selection in each cluster separately but does not select clusters. By combining the strengths of the two naïve algorithms, Ma and Huang (2007) proposed the CTGDR method that incorporates cluster structure into TGDR-based variable selection.

Let $\tau_1, \tau_2 \in [0,1]$ be two threshold parameters. In step 3 of the TGDR algorithm, define

$$f(\nu) = f_j^1(\nu) \times f_j^2(\nu),$$

where $f^1(\nu)$ and $f^2(\nu)$ are defined in the first naïve CTGDR and the second naïve CTGDR methods, respectively. Then $f^1(\nu)$ carries out cluster selection, while $f^2(\nu)$ carries out within-cluster gene selection. The combined CTGDR carries out feature selection both at the cluster level and within the cluster level. By allowing different threshold values of τ_1 and τ_2, more flexible results can be obtained.

Like the TGDR method, the properties of the CTGDR estimates are determined by the three tuning parameters

κ, τ_1, and τ_2. If τ_1 and τ_2 are both close to 1, then the estimate remains sparse for a relatively large κ but will become dense eventually. If τ_1 and τ_2 are both close to 0, then the estimate is dense for even a very small κ. With nonzero τ_1 and τ_2, the model with small to moderate κ usually has a small number of clusters and a small number of genes within each selected cluster.

Tuning parameter selection

Since the characteristics of the CTGDR estimates are determined by the three tuning parameters, κ, τ_1, and τ_2, the selection of these parameters should be well defined. Ma and Huang (2007) defined the cross-validated objective function, $CV(k)$, having chosen the tuning parameter κ that maximizes $CV(k)$ for any fixed (τ_1, τ_2), and obtained model features for different τ_1 and τ_2. Then, the parsimonious model with relatively large CV score has been chosen.

Results

Carcinoma of the lung has been the leading cause of cancer death in the United States and worldwide. Human lung carcinomas were classified by mRNA expression profiling, and distinct adenocarcinoma subclasses were revealed by Bhattacharjee *et al.* (2001). Hierarchical clustering was applied to recapitulate the distinctions between established historical classes of lung tumors and adenocarcinomas. Furthermore, the relationship between gene expression tumor classes and the survival times has been studied. However, it is very challenging to identify genes that have significant effects on survival time because of a large number of genes and relatively small sample sizes. To select significant variables from a large number of variables (e.g., genes) effectively, many methods have been developed, including LASSO and other regularization methods. We apply these methods to our data set in order to identify the significant genes for lung carcinomas.

A total of 203 snap-frozen lung tumors (n=186) and normal lung (n=17) specimens were sampled. The 203 specimens include histologically-defined lung adenocarcinomas (n=127), squamous cell lungcarcinomas (n=21), pulmonary carcinoids (n=20), small-cell lung cancer (n=6) cases, and normal lung (n=17) specimens. Other adenocarcinomas (n=12) were suspected to be extrapulmonary metastases based on clinical history. Using oligonucleotides, mRNA expression levels corresponding to 12,600 transcript sequences were analyzed from 186 lung tumor samples, including 139 adenocarcinomas resected from the lung. Among 12,600 transcript sequences, the 3312 most variable transcript sequences were selected by using a standard deviation threshold of 50 expression units. Only 125 adenocarcinoma samples were used for analysis due to availability of clinical data, such as the survival time. Genes were also standardized to have zero mean and unit variance. Applying the gap statistics, the number of clusters was selected as 15 (*K*-means) and 25 (Hierarchical), respectively, as shown in Fig. 1.

We obtained the result of model features with cross-validation-selected tuning parameters. As shown in Table 1, we compare the four methods, including two different CTGDR methods depending on the clustering methods. The TGDR method selects the largest number of genes while the LASSO method selects the smallest number of genes. The two different CTGDR methods select the similar number of genes. The whole list of genes selected by the four methods can be given on request.

Since a list of genes selected by four methods is not perfectly matched across methods, it is desirable to compare which genes are selected from all methods and how similar the selected genes are across methods. Table

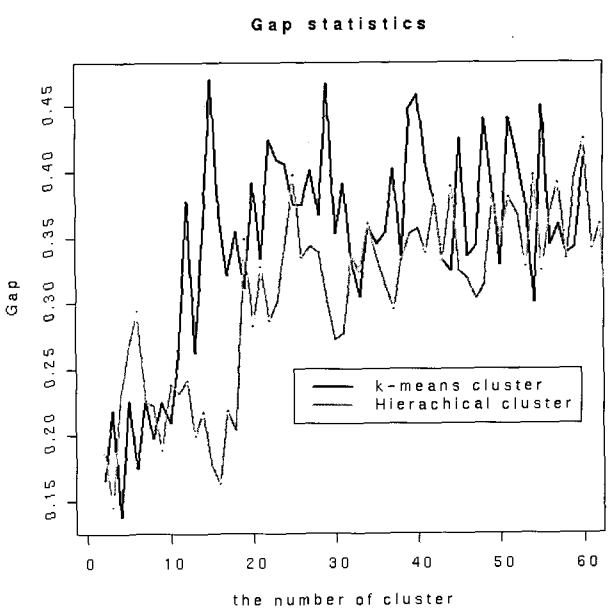

Fig. 1. Gap statistic as a function of number of clusters. Dark line: *K*-means clustering. Gray line: hierarchical clustering.

Table 1. Comparison of methods for gene selection

Method	Tuning	Non-zero Gene	Cluster
K-means-CTGDR	(τ_1, τ_2)=(1,1)	24	4
K-means-simple	-	3312	15
Hierarchical-CTGDR	(τ_1, τ_2)=(1,1)	23	3
Hierarchical-simple	-	3312	25
TGDR	τ=1 $\Delta\nu$=1e-04	31	-
LASSO	u=5.321 λ=23.728	6	-

2 displays a list ofgenes that are selected by more than two methods. Comparing the selected genes shown in Table 2, TGDR includes most of the genes selected by the other methods, while all of the genes selected by LASSO are included in those by TGDR, but only three genes overlapped with those selected by *K*-means CTGDR. Comparing the two different CTGDR methods, the 12 genes selected by *K*-means CTGDR are the same as those selected by the TGDR method, while only 5 genes selected by hierarchical CTGDR are the same as those selected by the TGDR method. Only four of the same genes are selected by both the *K*-means and hierarchical CTGDR methods. From these results, TGDR includes almost all genes selected by both the *K*-means and hierarchical CTDGR methods. This result implies that the selection of genes depends on the clustering information as well as the clustering methods. We can make a statistical inference about the association of genes with survival time from a list of genes selected across methods. For example, it is suspected that the seven genes selected by the three methods as displayed in Table 2 would be more strongly associated with survival time. Although TGDR and LASSO are effective regularization methods for variable selection, they do not consider cluster structures, whereas two different CTGDR methods incorporate cluster structures into TGDR-based variable selection. However, the two CTGDR methods select genes rather independently except for four overlapping genes. Moreover, when comparing the estimated coefficients of genes, the LASSO yields smaller values of the estimated coefficients than the other three methods. From the results in Table 2, it is not easy to know which method is the best for variable selection.

Table 2. A list of genes selected by more than two methods with the estimates of regression coefficients

Gene name	$\hat{\beta}$: coefficient of regressor			
Method	TGDR	Hierarchical-CTGDR	*K*-means CTGDR	LASSO
X37330_at	-0.0195*	-0.0218	-0.0271	
X31990_at	-0.0625	-0.1540	-0.1118	
X1707_g_at	-0.0376	-0.0157	-0.0077	
X35104_r_at		-0.0520	-0.0096	
X38833_at	-0.0516	-0.0734		-0.0008
X32623_at	-0.0630	-0.0089		
X41221_at	0.1120		0.1269	0.0949
X39079_at	0.1780		0.1142	0.0626
X40193_at	0.0908		0.1695	0.0421
X41332_at	0.0077		0.0373	
X34857_at	0.0111		0.0213	
X31477_at	0.0846		0.1021	
X32137_at	0.0544		0.1174	
X33453_at	-0.0031		0.0656	
X38791_at	-0.0815		-0.0778	
X41749_at	-0.0992			-0.0138
X40096_at	-0.0407			-0.0092

*The seven genes selected by three methods are written in bold.

Evaluation

To evaluate the prediction performance, we use the cross-validation-based approach, as suggested in Ma and Huang (2007), as follows:

First, we partition the data randomly into a training set of size n_1 and a test set of size n_2, with $n_1 + n_2 = n$ Then we compute the CTGDR estimates based on the training set only and compute a prediction index for the testing set using the training set estimates. To overcome the possibility of extreme prediction performance, repeat this process *B* (e.g., 100) times. For the censored survival data, we create two risk groups by dividing the testing set by a median risk score of the estimated linear risk scores $\hat{\beta}$Z. The estimates of regression coefficients of the selected genes are partly shown in Table 2. Then we compute the log-rank statistic for testing the equivalence of two survival curves of different risk groups. After repeating this procedure *B* times, we take the average of the log-rank statistics. If this value of the log-rank statistic is large, thenthe two risk groups are well separated, which implies that the prediction of the CTGDR estimates performs satisfactorily.

Alternatively, as suggested by a referee, we compute the different median risk score, which is obtained from the significant genes from the fitted Cox model in which only the selected genes are considered. In other words, the methods are evaluated with the risk scores of the different estimates from those shown in Table 2, which are displayed in Table 3. For example, for the LASSO method, the Cox model is fitted with six selected genes but only four significant genes of the six genes are involved in calculating the risk score and the log-rank statistic for testing the equivalence of the two risk groups.

Comparing the results of Table 2 and Table 3, only a few genes are overlapped within the two Tables. In addition, the estimates shown in Table 3 are much larger than those in Table 2, though the signs of the estimates are unchanged.

For the comparison of methods, we evaluate the prediction performance of methods using the two different log-rank statistics based on the 3-fold cross-validation with B=100 and 200 random partitions, respectively. As shown in Table 4, none of the methods provides the significant separation of the high- and low-risk groups by log-rank statistic calculated by the risk score from the estimates of each method, since the average values of the log-rank statistics are not large enough to be significant for testing

Table 3. A list of genes selected from the Cox model with the estimates of regression coefficients

Gene name	$\hat{\beta}$: coefficient of regressor			
Method	TGDR	Hierarchical-CTGDR	K-means CTGDR	LASSO
X41332_at	0.6998		0.3917	
X39079_at	0.5443			0.3615
X31990_at		-0.5089	-0.4563	
X36924_r_at		0.6796	0.8081	
X32623_at	-0.5843			
X32137_at	0.3935			
X38791_at	-0.6403			
X39758_at	0.5506			
X36070_at	0.3640			
X38392_at	0.5230			
X571_at	0.4785			
X38833_at		-0.3682		
X40075_at		0.6210		
X40095_at		-0.4504		
X39242_at		-0.4183		
X36838_at		-0.5661		
X41767_r_at			-0.5305	
X838_s_at			-0.4028	
X37678_at			-0.5437	
X37037_at			0.4935	
X34857_at			0.3538	
X41749_at				-0.3188
X40096_at				-0.4911
X40193_at				0.2994

Table 4. Evaluation result of four methods based on two different log-rank statistics

Method	The number of Non-zero Genes	Log-rank* (3-fold-CV)	The number of Significant Genes	Log-rank** (3-fold-CV)
K-means-CTGDR	24	0.5853	9	7.7179
Hierarchical-CTGDR	23	0.2914	7	4.8772
TGDR	31	0.4966	8	13.8748
LASSO	6	0.3350	4	6.0241

*log-rank statistic based on the risk score from the estimates of the selected genes with B=100

**log-rank statistic based on the risk score from the estimates of the significant genes in the Cox model with B=200

the equivalence of the two groups. However, the log-rank statistics show significant results for all methods when the risk scores are computed using the estimates of the significant genes in the fitted Cox model. Among those, TGDR has the largest value of the log-rank statistic, while the hierarchical CTGDR method has the smallest value of the log-rank statistic. The *K*-means CTGDR method and LASSO have similar values of the log-rank statistic. This result seems to be rather contradictory to what is expected because the two CTGDR methods do not show better performance than TGDR, even though these methods use more information about clustering structures than TGDR. According to the results shown in Table 4, the TGDR method performs better than the *K*-means and hierarchical CTGDR, while the LASSO method has slightly better predictive value than the hierarchical CTGDR method. However, the results from Table 4 can not be generalized because tuning parameter values and the number of clusters are not chosen over a variety of choices. For the generality of results, more performance should be implemented by considering a variety of parameters that can affect variable selection as well as evaluation of the performance.

Discussion

In this paper, we review four methods—LASSO, TGDR and two different CTGDR—to investigate which genes are significantly predictive of survival time using the Cox model. Since there is a large number of genes available, it is not easy to select susceptible genes for relatively small sample sizes. The LASSO method shrinks coefficients and produces some coefficients that are exactly zero, which identifies a small number of important genes. The TGDR method is also effective for variable selection using path information. However, these two methods do not use cluster structure, whereas the CTGDR method takes advantage of clustering structures.

Comparison of the four methods using an example of lung tumor data showed that the two different CTGDR methods yield more compact gene selection than TGDR, which includes almost all genes selected by two CTGDR methods, whereas LASSO provides a smaller subset of genes than other methods. Since the selected genes are not consistent for each method, it is difficult to determine which method provides the best selection of genes. The performance of methods is evaluated using 3-fold cross-validation based on two log-rank statistics. One of them is calculated using the risk score from the estimates of selected genes, while the other is calculated using the risk score from the estimates of the significant genes from the fitted Cox model. None of the methods provides any significant result in separating the high- and low-risk groups by the first log-rank statistic, whereas all methods yield significant results in separating the two risk groups with the second log-rank statistic. This is due to differences in the estimates of the coefficients depending on whether regularization is involved in the estimation process or not. The estimates for the Cox model are obtained without any regularized penalty, whereas the regularized estimates are obtained by using penalty to be selected from a large number of genes. Therefore, it seems to be more desirable

to estimate the effects of genes in the Cox model once the significant genes are selected using some regularization methods.

In addition, it would be more profitable to evaluate the methods by choosing from a variety of choices of tuning parameters, the number of clusters, and the validation statistics. Since the performance also depends on clustering information, it may be critical which information can be used for CTGDR. For example, if there are well-defined biological pathways, the proposed CTGDR method can makeuse of that information to select susceptible genes. Therefore, it would be desirable to extend this method with more information of pathways in future studies.

Acknowledgments

This work was supported by the National Research Laboratory Program of Korea Science and Engineering Foundation (M10500000126).

References

Bhattacharjee, A., Richards, W.G., Staunton, J., Li, C., Monti, S., Vasa, P., Ladd, C., Beheshti, J., Bueno, R., Gillette, M., Loda, M., Weber, G., Mark, E.J., Lander, E.S., Wong, W., Johnson, B.E., Golub, T.R., Sugarbaker, D.J., and Meyerson, M. (2001). Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc. Natl. Acad. Sci. USA* 98, 13790-13795.

Cox, D.R. (1972). Regression models and life tables (with discussion). *J.R. Stat. Soc. B.* 34, 187-202.

Friedman, J.H. and Popescue, B.E. (2004). Gradient directed regularization for linear regression and classification. *Technical report*, Department of Statistics, Stanford University. http://www-stat.stanford.edu/~jhf/PathSeeker.html.

Gui, J. and Li, H. (2005). Threshold gradient descent method for censored data regression with applications in pharmacogenomics. *Proc. Pac. Symp. Biocomput.* 10, 272-283.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. Chapman and Hall.

Ma, S. and Huang, J. (2007). Clustering threshold gradient descent regularization: with applications to microarray studies. *Bioinformatics* 23, 466-472.

Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E.S., and Golub, T.R. (1999). Interpreting patterns of gene expression with self-organizing maps: Methods and application to hermatopoetic differentiation. *Proc. Natl. Acad. Sci. USA* 96, 2907-2912.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J.R. Stat. Soc., B.* 58, 267-288.

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