

A novel Log penalty in a path seeking scheme for biomarker selection

Sai Wang^a, Hui Zhang^a, Hua Chai^a and Yong Liang^{a,b,*}

^a*Faculty of Information Technology, Macau University of Science and Technology, Macau, China*

^b*State Key Laboratory of Quality Research in Chinese Medicines, Macau University of Science and Technology, Macau, China*

Abstract. Biomarker selection or feature selection from survival data is a topic of considerable interest. Recently various survival analysis approaches for biomarker selection have been developed; however, there are growing challenges to currently methods for handling high-dimensional and low-sample problem. We propose a novel Log-sum regularization estimator within accelerated failure time (AFT) for predicting cancer patient survival time with a few biomarkers. This approach is implemented in path seeking algorithm to speed up solving the Log-sum penalty. Additionally, the control parameter of Log-sum penalty is modified by Bayesian information criterion (BIC). The results indicate that our proposed approach is able to achieve good performance in both simulated and real datasets with other ℓ_1 type regularization methods for biomarker selection.

Keywords: Biomarker selection, Log-sum regularization, path seeking scheme, Bayesian information criterion

1. Introduction

Biomarker selection or feature selection from survival data is a topic of considerable interest. The regularization methods are group of feature selection methods that embed different penalized methods in the learning procedure into a single process. In this way, the regularization methods could reduce the overfitting problem. The ℓ_0 -norm is a total number of non-zero elements in a vector, but it faces the problem of combinatory optimization, i.e., ℓ_0 -norm is too complex and almost impossible to solve. Thus, alternative sparsity-promoting functions which are more computationally efficient in finding the sparse solution are desirable. The most popular alternative is to replace the ℓ_0 -norm with the ℓ_1 -norm (or Lasso) penalty function that is the least absolute shrinkage and selection operator [1]. There are also some ℓ_1 type regularizations can also take place of ℓ_0 -norm, like smoothly clipped absolute deviation (or SCAD) [2], minimax concave penalty (or MCP) [3], group lasso [4]. Moreover, the general sparse representation method is to solve a linear representation system with the ℓ_p -norm minimization problem, especially $p = 0.1, 1/2, 2/3$ or 0.9 [5,6]. Recently, Candès et al. [7] proposed the Log-sum penalty that approximated ℓ_0 -norm much better than other penalties by reweighting the ℓ_1 -norm of transformed object. As shown in Fig. 1, these five states of art penalty methods satisfy properties of sparsity and continuity. Especially, the Log-sum penalty is sparser than other four penalty methods. Unlike Lasso and Elastic net biased estimators, the $\ell_{1/2}$, SCAD and Log-sum penalties have unbiasedness, i.e., they can

*Corresponding author: Yong Liang, Faculty of Information Technology, Macau University of Science and Technology, Taipa, Macau 999078, China. E-mail: yliang@must.edu.mo.

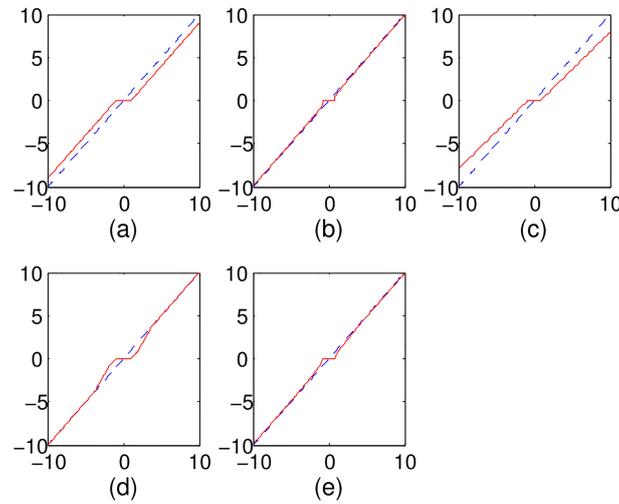


Fig. 1. Various penalized function for orthonormal design: (a) Lasso, (b) $\ell_{1/2}$, (c) Elastic net, (d) SCAD, (e) Log-sum.

easily construct unbiased estimators when the coefficient is large. Then, Chartrand and Yin [8] used this iteratively reweighted ℓ_1 minimization algorithm in sparse signal recovery. Xia et al. [9] proposed multiple linear regression with Log-sum penalty in a thresholding representation theory for drug discovery. We take the advantage of the Log-sum penalty in strong sparsity and employ it to select the key factors without any prior information in survival data.

For survival analysis, there are mainly two types of survival model, namely Cox model or Cox proportional hazards model [10] and accelerated failure time (AFT) model [11]. The Cox model usually is used for predicting the hazard rate of a disease, whereas the AFT model is available to estimate survival time of the patients with simply regressing the exponential over the key risk predictors [12]. Besides, the physical interpretation of AFT model is similar with standard regression, so the AFT model comes out as an attractive alternative to the Cox proportional hazard model for censored failure time data [12]. Furthermore, the log-linear form of AFT model increases its robustness to the model misspecification and yield narrower confidence interval for regression coefficients [13]. After the penalized Cox model with ℓ_1 -norm [14], Datta et al. [15] combined AFT model with ℓ_1 -norm to predict failure time outcomes. Other penalized AFT models aim to obtain more accurate and sparse predictor for survival analysis, such as AFT via bridge penalization [16], $\ell_{1/2}$ -norm AFT model [17], etc. In this article, we employ path seeking scheme to accelerate solving the Log-sum penalty for predicting patient survival time with fewer biomarkers.

The rest of this article is organized as follows: Section 2 introduces the Log-sum penalized AFT model. In Section 3, we implement our proposed novel Log penalty in path seeking scheme. Finally, we discuss the experimental results in Section 4 and make some conclusions in Section 5.

2. Log-sum penalized AFT model

Suppose the survival data have h patients $(\tau_i, \delta_i, \dots, X_i)_{i=1}^h$, where the sample vector of a survival times $\mathbf{T} = (\tau_1, \tau_2, \dots, \tau_h)^T$ and $\tau_i = \min(t_i, c_i)$, here t_i is the true survival time and c_i is the time to the first censoring event (e.g., study conclusion, date of final follow up) for each sample i . The δ indicates

the censoring time, i.e.,

$$\delta_i = \begin{cases} 0, & \text{the right censoring time} \\ 1, & \text{the completed time} \end{cases}.$$

X_i denotes the gene expression data of i -th patient, i.e., $X_i = (x_{i1}, x_{i2}, \dots, x_{ik})$, where k is the number of genes.

The accelerated failure time (AFT) model is used to define the survival time τ_i as follows:

$$\tau_i = \exp(\beta_0 + X_i \beta^T + \varepsilon_i) \quad (1)$$

where $\beta \subseteq \mathbb{R}^k$ is the coefficient vector of k variables, β_0 is the intercept, and $\varepsilon_i \sim N(0, 1)$ is independent random error. In this article, we employ the mean imputation method [15] that converts the censoring survival time τ_i to the estimated survival time $G(\tau_i)$ as the following estimated function:

$$G(\tau_i) = \delta_i \log(\tau_i) + (1 - \delta_i) \{\widehat{S}(\tau_i)\}^{-1} \sum_{t_{(r)} > \tau_i} \log(t_{(r)} \Delta \widehat{S}(t_{(r)})) \quad (2)$$

where r is the amount of individuals at risk of failing just before time $t_{(i)}$ that are different censored survival times in an ascending order, and $\Delta \widehat{S}(t_{(r)})$ is the step of Kaplan-Meier estimator \widehat{S} at time $t_{(r)}$ [18]. Then, we can directly using AFT model via standard least squares approach to minimize the loss function $L(\beta)$ as follows:

$$L(\beta) = \frac{1}{h} \sum_{i=1}^h \left(y_i - \sum_{j=0}^k \beta_j x_{ij} \right)^2 \quad (3)$$

where y_i is replaced by the estimated survival time $G(\tau_i)$ in Eq. (2), i.e., the survival times τ_i logarithmically transformed into y_i .

3. Implementation of Log-sum penalized AFT model

The regularization methods are used to reduce the overfitting problem of learning procedure through adding the penalty term, therefore the general regularization can be modeled as:

$$\widehat{\beta}(\lambda) = \underset{\beta}{\operatorname{argmin}} \{L(\beta) + \lambda P(\beta)\} \quad (4)$$

where $\beta \subseteq \mathbb{R}^k$ is the coefficient of covariate, $\lambda > 0$ is a control parameter, $L(\beta)$ represents the loss term and $P(\beta)$ is the penalty term. Larger values of λ exert higher penalties on regression coefficients, resulting on inclusion of fewer variables in the model and vice versa. The generalized cross-validation [19] has been widely used for given an appropriate value of the control parameter. Huang et al. [20] used a modified Akaike's information criterion (AIC) for choosing tuning parameter. Wang and Song [21] used Bayesian information criterion (BIC) for tuning parameter selection under AFT model with adaptive Lasso. Friedman [22] obtained the control parameter by solving the component ratios of the gradient of the loss function and regularization term that is called generalized path seeking scheme. This scheme is much faster than general convex optimizers for squared-error loss.

For the regularization term $P(\beta)$, many penalties are proposed to bridge the gap between the ℓ_0 and ℓ_1 minimization. Such a Log-sum penalty function was originally introduced in [23] for basis selection

which indicates that Log-sum based methods present uniform superiority over the conventional ℓ_1 -type methods. The Log-sum sparsity-encouraging functional for survival analysis leads to:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left\{ L(\beta) + \lambda \sum_{j=1}^k \log(|\beta_j| + \xi) \right\} \quad (5)$$

where $\xi > 0$ is a positive parameter to ensure that the function is well-defined. Especially, the Log-sum penalty function behaves like the ℓ_0 -norm when $\xi \rightarrow 0$ [24]. In this article, we employ the path seeking method [22] to solve the Log-sum penalized AFT model through constructing a path directly and successively in parameter space. Let v measure length along the path and the step size $\Delta v > 0$ can be calculated by

$$\frac{L(\hat{\beta}(v)) - L(\hat{\beta}(v + \Delta v))}{L(\hat{\beta}(v))} = 0.01 \quad (6)$$

Define

$$\begin{aligned} \varphi_j(v) &= - \left[\frac{\partial L(\beta)}{\partial \beta_j} \right]_{\beta = \hat{\beta}(v)} \\ &= - \left[\frac{\partial \frac{1}{h} \sum_{i=1}^h (y_i - \sum_{j=0}^k \beta_j x_{ij})^2}{\partial \beta_j} \right]_{\beta = \hat{\beta}(v)} \end{aligned} \quad (7)$$

$$\begin{aligned} &= \left[\frac{2}{h} \sum_{i=1}^h x_{ij} \left(y_i - \sum_{j=0}^k \beta_j x_{ij} \right) \right]_{\beta = \hat{\beta}(v)} \\ \phi_j(v) &= \left[\frac{\partial \sum_{j=1}^k \log(|\beta_j| + \xi)}{\partial |\beta_j|} \right]_{\beta = \hat{\beta}(v)} \\ &= \left[\frac{\partial \log(|\beta_j| + \xi)}{\partial |\beta_j|} \right]_{\beta = \hat{\beta}(v)} \\ &= \left[\frac{1}{|\beta_j| + \xi} \right]_{\beta = \hat{\beta}(v)} \end{aligned} \quad (8)$$

$$\lambda_j(v) = \frac{\varphi_j(v)}{\phi_j(v)} \quad (9)$$

where $\lambda_j(v)$ is the value of λ in Eq. (4) corresponding to v , and is also the ratio of loss function gradient $\varphi_j(v)$ for $L(\beta)$ in Eq. (3) and penalty function gradient $\phi_j(v)$ with respect to $|\beta_j|$. Without estimating λ , this path seeking scheme can accelerate solving the Log-sum penalty. Besides, ξ is chosen by ten-fold cross-validation. The details of the achievement for Log-sum penalty are represented in Algorithm 1.

At first, we initialize the path, and then compute the vector $\lambda(v)$ by Eqs (7)–(9) in each step. Subsequently, the non-zero coefficients $\hat{\beta}_j(v)$ are recognized. Those $\hat{\beta}_j(v)$ have a sign opposite to that of their corresponding $\lambda_j(v)$. Generally there are non the coefficient corresponding to the largest component of $\lambda_j(v)$ in absolute value is selected. If one or more $\lambda_j(v) \cdot \hat{\beta}_j(v) < 0$, then the coefficient with

Algorithm 1 The algorithm of Log-sum penalty

```

1. Initialize:  $v = 0, \{\widehat{\beta}_j(0) = 0\}_1^k$ 
2. repeat
3.   Compute  $\{\lambda_j(v)\}_1^k$ 
4.    $S = \{j | \lambda_j(v) \cdot \widehat{\beta}_j(v) < 0\}$ 
5.   if  $S = \text{empty}$  then
6.      $j^* = \operatorname{argmax}_j |\lambda_j(v)|$ 
7.   else
8.      $j^* = \operatorname{argmax}_{j \in S} |\lambda_j(v)|$ 
9.   end if
10.   $\widehat{\beta}_{j^*}(v + \Delta v) = \widehat{\beta}_{j^*}(v) + \Delta v \cdot \operatorname{sign}(\lambda_{j^*}(v))$ 
11.   $\{\widehat{\beta}_j(v + \Delta v) = \widehat{\beta}_j(v)\}_{j \neq j^*}$ 
12.   $v \leftarrow v + \Delta v$ 
13. until  $\lambda_j(v) = 0$ 

```

corresponding large $|\lambda_j(v)|$ within this subset is instead selected. The selected coefficient $\widehat{\beta}_{j^*}(v)$ is then incriminated through a small amount in the direction of the sign of its correspond $\lambda_{j^*}(v)$ with all other coefficient residual unchanged, producing the solution for the next path point $v + \Delta v$. Iterations continue until all components of $\lambda(v)$ are zero.

4. Numerical experiments

4.1. Simulated datasets

In order to simulate the high-dimensional and low-sample property of gene expression data, we assumed that 20 nonzero factors among $k = 2000$ variables with different fraction and sample size $h = 90, 300$ respectively based on the following model:

$$Y = \sum_{u=1}^{20} \beta_u X_u + \sigma \varepsilon \quad (10)$$

where Y denotes the vector of survival times logarithmically transformed $y_i = \log(\tau_i)$ in Eq. (3) without censored data, i.e., $Y = (y_1, y_2, \dots, y_h)$, ε is an independent random noise that is generated from a normal distribution $N(0, 1)$, σ controls the noise strength and the coefficients of relevant features are specified as

$$\beta = (\underbrace{2, \dots, 2}_5, -2, 1.5, -1.7, 2.5, -1.8, \underbrace{4, \dots, 4}_{10}, \underbrace{0, \dots, 0}_{1980})$$

The \mathbf{X} value is simulated from an array $tmp_{i0, \dots, ik}$, ($i = 1, \dots, h$) of independent standard normal distribution:

$$x_{ij} = \sqrt{\varrho} \times tmp_{i0} + \sqrt{(1-\varrho)} \times tmp_{ij} \quad (11)$$

where the correlation coefficient ϱ are 0.1 and 0.3 respectively in our experiment.

Additionally, the both *sensitivity* and *specificity* for each procedure are calculated as follows:

$$\text{sensitivity} = \frac{\# \text{ correctly selected genes}}{\# \text{ non-zero in } \beta} \quad (12)$$

Table 1
The results of different penalized methods in simulated data

σ	ϱ	Penalty	$h = 90$				$h = 300$			
			Sensitivity	Specificity	MSE	CI	Sensitivity	Specificity	MSE	CI
1.0	0.1	Lasso	0.758	0.896	87.924	0.808	0.850	0.971	74.175	0.908
		$\ell_{1/2}$	0.856	0.908	57.750	0.863	0.900	0.978	29.825	0.913
		Elastic net	0.782	0.731	149.340	0.638	0.800	0.754	117.393	0.720
		SCAD	0.845	0.891	67.254	0.792	0.900	0.977	43.502	0.872
		Log-sum	0.902	0.953	54.615	0.871	0.950	0.986	23.845	0.928
	0.3	Lasso	0.684	0.786	98.183	0.782	0.762	0.864	64.536	0.858
		$\ell_{1/2}$	0.793	0.894	62.786	0.809	0.850	0.935	59.747	0.872
		Elastic net	0.912	0.726	174.570	0.800	0.999	0.757	141.602	0.722
		SCAD	0.760	0.882	74.009	0.765	0.800	0.945	47.681	0.885
		Log-sum	0.841	0.931	58.892	0.847	0.872	0.978	46.893	0.918
1.5	0.1	Lasso	0.721	0.860	71.004	0.643	0.800	0.967	54.601	0.762
		$\ell_{1/2}$	0.840	0.903	63.549	0.725	0.910	0.972	40.219	0.887
		Elastic net	0.750	0.683	198.627	0.569	0.850	0.705	179.118	0.641
		SCAD	0.767	0.889	69.073	0.703	0.850	0.965	51.669	0.847
		Log-sum	0.880	0.935	62.720	0.804	0.950	0.982	38.358	0.915
	0.3	Lasso	0.617	0.756	119.453	0.691	0.650	0.861	95.235	0.811
		$\ell_{1/2}$	0.750	0.857	82.944	0.714	0.800	0.928	60.812	0.849
		Elastic net	0.860	0.605	213.905	0.571	0.959	0.621	185.917	0.645
		SCAD	0.751	0.879	86.892	0.609	0.800	0.937	68.816	0.638
		Log-sum	0.800	0.917	65.092	0.784	0.850	0.974	48.973	0.890

$$\begin{aligned}
 &= \frac{\# \text{ correctly selected genes}}{20} \\
 \text{specificity} &= \frac{\# \text{ correctly rejected genes}}{\# \text{ zero in } \beta} \\
 &= \frac{\# \text{ correctly rejected genes}}{1980}
 \end{aligned} \tag{13}$$

The optimal combination of ξ is selected under ten-fold cross-validation by minimizing the Bayesian information criterion (BIC) defined as

$$\text{BIC}_\xi = -\frac{2}{h} \log(\text{mse}(\beta)) - \frac{\log(h)}{2} df_\xi \tag{14}$$

where h is the total number of observations; df_ξ is the number of nonzero parameters; and $\text{mse}(\beta)$ measures the mean square error that is defined by

$$\text{mse}(\beta) = \frac{1}{h} \sum_{i=1}^h (\tau_i - \hat{\tau}_i)^2 \tag{15}$$

where the predicted value $\hat{\tau}_i = \exp(\sum_{j=0}^k \beta_j x_{ij})$. In our simulations and application, the optimal ξ is searched on grid points.

We also employ the concordance index (CI) to evaluate the predictive accuracy of survival models. CI or c-index can be interpreted as the fraction of all pairs of subjects whose predicted survival times are correctly ordered among all subjects that can actually be ordered. Therefore, it can be written as:

$$ci(\beta) = \frac{\sum_i \sum_j 1(\hat{\tau}_i < \hat{\tau}_j \text{ and } \delta_i = 1)}{\sum_i \sum_j 1(\tau_i < \tau_j \text{ and } \delta_i = 1)} \tag{16}$$

Table 2
The results in the GSE22210

Penalty	# Selected genes	MSE	CI
Lasso	46	21.023	0.776
$\ell_{1/2}$	23	25.271	0.783
Elastic net	159	33.331	0.809
SCAD	33	22.754	0.791
Log-sum	15	16.814	0.821

Table 3
The selected genes of different penalized methods in the GSE22210

	Lasso	$\ell_{1/2}$	Elastic net	SCAD	Log-sum
1	XIST	IL1B	SERPINB2	IL1B	XIST
2	LAT	XIST	XIST	XIST	LAT
3	IL1B	HLA-DQA2	IMPACT	CCND1	DNASE1L1
4	DNASE1L1	TGFA	IL1B	HLA-DQA2	IL1B
5	NFKB1	CDKN1A	LAT	NFKB1	BCL2L2
6	HDAC9	GNMT	CCND1	PTHR1	MEST
7	BCL2L2	LAT	NFKB1	GNMT	NFKB1
8	ESR2	BCL2L2	TGFA	LAT	DAB2IP
9	AFP	HDAC9	HLA-DQA2	CD44	ESR2
10	LAMC1	CD44	RASGRF1	ESR2	APC

The simulated experiments are repeated 100 times. From Table 1, we can conclude that the Log-sum penalty using the path seeking algorithm can achieve lower MSE with higher CI than other penalties. Furthermore, this Log-sum penalty results in higher sensitivity for identifying correct genes compared to the other four algorithms. With increasing sample size, the performance of Log-sum penalty is better. For example, when $\sigma = 1.0$, $\rho = 0.1$, the performance of Log-sum in $h = 300$ has lower MSE with higher CI than in $h = 90$. The Elastic Net with ℓ_2 -norm selected the largest number of genes within the synthetic data with poor performance, e.g., when $\sigma = 1.0$, $\rho = 0.3$, $h = 300$ Elastic Net nearly selected 20 non-zero coefficients (i.e., 20×0.999), but it selected 481 irrelevant coefficients (i.e., 1980×0.243). With the correlation ρ increasing among genes for various noise levels σ , the ℓ_1 -norm (Lasso) cannot distinguish the key genes very well, while other ℓ_1 type penalties performed effectively, especially the Log-sum penalty.

4.2. Real datasets

To further demonstrate the performance of these regularization methods, we compare our proposed method with other four penalties on GSE22210 microarray expression data from NCBI's gene expression omnibus (GEO). This breast cancer dataset includes 1,452 genes and 167 samples [25]. We divide the data set at random two-thirds samples (117 samples) are training set and the remainders (50 samples) are used to test. Table 2 shows that the Log-sum penalty achieves best predicting survival time just with fewer genes than other ℓ_1 type regularization methods.

As seen from the Table 3, some genes are selected by all methods such as XIST, LAT and IL1B. Missing from *XIST* RNA, the X chromosome causes the basal-like subtype of invasive breast cancer [26]. Furthermore, *LAT* is short for Linker for Activation of T cells that plays a crucial role in the TCR-mediated signaling pathways. The adoptive transfer of T cells appears to be a promising new treatment for various type-s of cancer [27]. Collado-Hidalgo et al. [28] provided evidence that polymorphisms in *IL1B* increase the production of proinflammatory cytokines triggered by the treatment, which subsequently affects persistent fatigue in the aftermath of breast carcinoma. There are some unique genes

selected by Log-sum, such as *MEST*, *DAB2IP*, *APC*, etc. *MEST* is also known as paternally expressed gene 1 (*PEG1*) that are often detected in invasive breast carcinomas [29]. The *DAB2IP* as a bona fide tumor suppressor that frequently silenced by promoter methylation in aggressive human tumors [30]. Furthermore, aberrant methylation of the *APC* gene is frequent in breast cancers [31].

5. Conclusion

In this paper, we propose a novel Log-sum regularization estimator with the AFT model in the path seeking scheme. Comparing with other ℓ_1 type penalties, the results in both simulated and real datasets indicate that our proposed Log-sum penalty can effectively predict patient survival time with fewer biomarkers. Thus, we believe it will be an effective tool for gene selection on high dimensional biological data.

Acknowledgments

The authors thank Dr. Zi-Yi Yang for excellent technical assistance. This work was supported by the Macau Science and Technology Develop Funds (Grant no. 003/2016/AFJ) of Macao SAR of China and China NSFC project (Contract no. 61661166011).

Conflict of interest

None to report.

References

- [1] Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B (Methodological)*. 1996; 267-288.
- [2] Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*. 2001; 96(456): 1348-1360.
- [3] Zhang CH. Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*. 2010; 38(2): 894-942.
- [4] Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2006; 68(1): 49-67.
- [5] Lyu Q, Lin Z, She Y, Zhang C. A comparison of typical Lpminimization algorithms. *Neurocomputing*. 2013; 119: 413-424.
- [6] Xu Z, Chang X, Xu F, Zhang H. L1/2regularization: A thresholding representation theory and a fast solver. *IEEE Transactions on neural networks and learning systems*. 2012; 23(7): 1013-1027.
- [7] Candes EJ, Wakin MB, Boyd SP. Enhancing sparsity by reweighted L1 minimization. *Journal of Fourier Analysis and Applications*. 2008; 14(5): 877-905.
- [8] Chartrand R, Yin W. Iteratively reweighted algorithms for compressive sensing. In: *Acoustics, speech and signal processing, 2008; ICASSP 2008. IEEE international conference on. IEEE; 2008. pp. 3869-3872.*
- [9] Xia LY, Wang YW, Meng DY, Yao XJ, Chai H, Liang Y. Descriptor Selection via Log-Sum Regularization for the Biological Activities of Chemical Structure. *International Journal of Molecular Sciences*. 2017; 19(1): 30.
- [10] Cox DR. Regression models and life-tables. In: *Breakthroughs in statistics*. Springer; 1992; p. 527-541.
- [11] Kalbfleisch JD, Prentice RL. *The statistical analysis of failure time data*. vol. 360; John Wiley & Sons; 2011.
- [12] Wei LJ. The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Statistics in Medicine*. 1992; 11(14-15): 1871-1879.

- [13] Hutton J, Monaghan P. Choice of parametric accelerated life and proportional hazards models for survival data: asymptotic results. *Lifetime Data Analysis*. 2002; 8(4): 375-393.
- [14] Tibshirani R. The lasso method for variable selection in the Cox model. *Statistics in Medicine*. 1997; 16(4): 385-395.
- [15] Datta S, Le-Rademacher J, Datta S. Predicting patient survival from microarray data by accelerated failure time modeling using partial least squares and LASSO. *Biometrics*. 2007; 63(1): 259-271.
- [16] Huang J, Ma S. Variable selection in the accelerated failure time model via the bridge method. *Lifetime Data Analysis*. 2010; 16(2): 176-195.
- [17] Chai H, Liang Y, Liu XY. The L1/2 regularization approach for survival analysis in the accelerated failure time model. *Computers in Biology and Medicine*. 2015; 64: 283-290.
- [18] Datta S. Estimating the mean life time using right censored data. *Statistical Methodology*. 2005; 2(1): 65-69.
- [19] Craven P, Wahba G. Smoothing noisy data with spline functions. *Numerische Mathematik*. 1978; 31(4): 377-403.
- [20] Huang J, Ma S, Xie H. Regularized Estimation in the Accelerated Failure Time Model with High-Dimensional Covariates. *Biometrics*. 2006; 62(3): 813-820.
- [21] Wang X, Song L. Adaptive Lasso variable selection for the accelerated failure models. *Communications in Statistics-Theory and Methods*. 2011; 40(24): 4372-4386.
- [22] Friedman JH. Fast sparse regression and classification. *International Journal of Forecasting*. 2012; 28(3): 722-738.
- [23] Coifman RR, Wickerhauser MV. Entropy-based algorithms for best basis selection. *IEEE Transactions on information theory*. 1992; 38(2): 713-718.
- [24] Rao BD, Kreutz-Delgado K. An affine scaling methodology for best basis selection. *IEEE Transactions on Signal Processing*. 1999; 47(1): 187-200.
- [25] Holm K, Hegardt C, Staaf J, Vallon-Christersson J, Jönsson G, Olsson H, et al. Molecular subtypes of breast cancer are associated with characteristic DNA methylation patterns. *Breast Cancer Research*. 2010; 12(3): R36.
- [26] Richardson AL, Wang ZC, De Nicolo A, Lu X, Brown M, Miron A, et al. X chromosomal abnormalities in basal-like human breast cancer. *Cancer Cell*. 2006; 9(2): 121-132.
- [27] June CH. Adoptive T cell therapy for cancer in the clinic. *The Journal of Clinical Investigation*. 2007; 117(6): 1466-1476.
- [28] Collado-Hidalgo A, Bower JE, Ganz PA, Irwin MR, Cole SW. Cytokine gene polymorphisms and fatigue in breast cancer survivors: Early findings. *Brain, Behavior, and Immunity*. 2008; 22(8): 1197-1200.
- [29] Pedersen IS, Dervan PA, Broderick D, Harrison M, Miller N, Delany E, et al. Frequent loss of imprinting of PEG1/MEST in invasive breast cancer. *Cancer Research*. 1999; 59(21): 5449-5451.
- [30] Di Minin G, Bellazzo A, Dal Ferro M, Chiaruttini G, Nuzzo S, Biciato S, et al. Mutant p53 reprograms TNF signaling in cancer cells through interaction with the tumor suppressor DAB2IP. *Molecular Cell*. 2014; 56(5): 617-629.
- [31] Virmani AK, Rathi A, Sathyanarayana UG, Padar A, Huang CX, Cunningham HT, et al. Aberrant methylation of the adenomatous polyposis coli (APC) gene promoter 1A in breast and lung carcinomas. *Clinical Cancer Research*. 2001; 7(7): 1998-2004.