Thailand Statistician
July 2019; 17(2): 212-222
http://statassoc.or.th
Contributed paper

Penalized Linear Regression Methods where the Predictors Have Grouping Effect

Kanyalin Jiratchayut*[a] and Chinnaphong Bumrungsup [b]
[a] Faculty of Science and Arts, Burapha University, Chanthaburi Campus, Tha-Mai, Chanthaburi, 22170, Thailand.
[b] Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, Khlong Luang, Pathum Thani, 12120, Thailand.
*Corresponding author; e-mail: kanyalinn@gmail.com

Received: 5 February 2018
Revised: 9 October 2018
Accepted: 14 December 2018

Abstract

The aim of this paper is to study the performance of four different penalized linear regression methods: elastic net, adaptive elastic net, L1CP, and SCAD-L_2. Simulation studies show that the adaptive elastic net performs best in variable selection and parameter estimation, while the SCAD-L_2 has prediction accuracy better than the other methods. When sample size is large, the L1CP has a prediction performance close to the prediction accuracy of the SCAD-L_2.

Keywords: Adaptive elastic net, correlation based penalty, elastic net, SCAD-L_2, variable selection.

1. Introduction

Penalized regression methods are developed to solve the regression analysis when multicollinearity problem exists. The penalized regression is developed from least squares method with penalty function to discover relevant explanatory factors and to get higher prediction accuracy in linear regression. The examples of penalized regression are ridge regression (Hoerl and Kennard 1970a, b), the lasso (Tibshirani 1996), and elastic net (Zou and Hastie 2005). The elastic net is developed to solve some drawback of the lasso. It can select groups of correlated predictors. The elastic net penalty is a combination of the lasso (L_1) and ridge (L_2) penalties. Consider a linear regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{y}$ is an $n \times 1$ vector of response variable, $\mathbf{X}$ is an $n \times p$ matrix of predictor variables, $\boldsymbol{\beta}$ is a $p \times 1$ vector of parameter of regression coefficients, $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of random errors, p is the number of predictors, and n is the number of observations. The errors are assumed to be independent identically normally distributed random variable with mean 0 and finite variance

σ^2. Without loss of generality, we assume the response is centered and the predictors are standardized, so the intercept is not included in the regression function. The naïve elastic net estimator is

$$\hat{\boldsymbol{\beta}}_{\text{elastic net}} = \arg\min_{\boldsymbol{\beta}} \left[\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda_2 \|\boldsymbol{\beta}\|_2^2 + \lambda_1 \|\boldsymbol{\beta}\|_1 \right], \tag{2}$$

where $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$ are the penalty parameters, $\lambda = \lambda_1 + \lambda_2$, and $\alpha = \lambda_2 / (\lambda_1 + \lambda_2)$ where $\alpha \in (0,1)$. The L_1 part of the elastic net performs automatic variable selection, while the L_2 part stabilizes the solution parts and, hence, improves the prediction. By introducing the L_2 norm penalty, the elastic net encourage grouping effect-group of highly correlated predictors tend to be in or out of the model together once one predictor among them is selected.

Similar to the elastic net, other penalized linear regression methods which have the ability to achieve grouping effect have been proposed, such as adaptive elastic net methods (Zou and Zhang 2009, and Ghosh 2011), penalized regression combining the L_1 norm and a correlation based penalty (L1CP) (Anbari and Mkhadri 2014), and SCAD-L_2 (Zeng and Xie 2014).

Zou and Zhang (2009), and Ghosh (2011) proposed two adaptive elastic net estimators. They added the adaptive weight into the L_1 penalty of the elastic net. Two adaptive elastic net approaches are different in their adaptive weights. Zou and Zhang (2009) construct the adaptive weight by using the elastic net estimator, whereas Ghosh (2011) use the least squares estimator to construct the adaptive weight. The study of Jiratchayut (2015), and Jiratchayut and Bumrungsup (2015a) revealed that the adaptive elastic net performs best in estimation accuracy and variable selection performance when the adaptive weight is constructed by using the elastic net estimator. The naïve adaptive elastic net estimator proposed by Zou and Zhang (2009) is defined as follows:

$$\hat{\boldsymbol{\beta}}_{\text{adaptive elastic net}} = \arg\min_{\boldsymbol{\beta}} \left[\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda_2 \|\boldsymbol{\beta}\|_2^2 + \lambda_1 \sum_{j=1}^{p} \hat{w}_j \left|\beta_j\right| \right]. \tag{3}$$

Let γ be a positive constant, the adaptive weight $\hat{w}_j = \left(\left|\hat{\beta}_{j(\text{elastic net})}\right|\right)^{-\gamma}$, $j = 1, \dots, p$.

Anbari and Mkhadri (2014) proposed the penalized regression combining the L_1 norm and a correlation based penalty (L1CP). They revised the L_2 penalty of the elastic net by using the correlation based penalty proposed by Tutz and Ulbricht (2009). The naïve L1CP estimator is

$$\hat{\boldsymbol{\beta}}_{\text{L1CP}} = \arg\min_{\boldsymbol{\beta}} \left[\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda_1 \|\boldsymbol{\beta}\|_1 + \lambda_2 P_C(\beta) \right], \tag{4}$$

where $P_C(\beta) = \sum_{j=1}^{p-1} \sum_{i>j} \left\{ \frac{(\beta_i - \beta_j)^2}{1 - \rho_{ij}} + \frac{(\beta_i + \beta_j)^2}{1 + \rho_{ij}} \right\}$, $\lambda_1 > 0$ and $\lambda_2 > 0$.

The SCAD-L_2 approach was proposed by Zeng and Xie (2014). They replaced the L_1 penalty of elastic net by the SCAD (smoothly clipped absolute deviation) method (Fan and Li 2001), so the penalty function of SCAD-L_2 is a combination of the SCAD function (Fan and Li 2001) and the L_2 penalty. The naïve SCAD-L_2 estimator is

$$\hat{\boldsymbol{\beta}}_{\text{SCAD-}L_2} = \arg\min_{\boldsymbol{\beta}} \left\{ \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \sum_{j=1}^{p} f_{\lambda_1}(\beta_j) + \lambda_2 \|\boldsymbol{\beta}\|_2^2 \right\}, \tag{5}$$

where the SCAD function is defined as

$$f_\lambda(\theta) = \begin{cases} \lambda|\theta| & \text{if } 0 \le |\theta| < \lambda, \\ -\dfrac{\theta^2 - 2a\lambda|\theta| + \lambda^2}{2(a-1)} & \text{if } \lambda \le |\theta| < a\lambda, \\ \dfrac{(a+1)\lambda^2}{2} & \text{otherwise,} \end{cases} \tag{6}$$

with a constant $a > 2.$

The elastic net, adaptive elastic net, L1CP, and SCAD-L_2 have the ability to do group selection in a similar way, whereas there is some difference in L_1-part and L_2-part of these approaches. There is no comparative study among the adaptive elastic net, L1CP, and SCAD-L_2 methods.

In this paper, we study the performance of the elastic net, adaptive elastic net, L1CP, and SCAD-L_2 approaches. We limit our attention to full rank model $(p < n).$ This paper is organized as follows. Section 2 describes the simulation method, comparative study and tuning parameters selection, and decision criterion. Section 3 presents the results and discussion. In Section 4, we illustrate our study using real datasets. Conclusions are presented in Section 5.

2. Materials and Methods

2.1. Simulation data

The datasets are simulated by the simulation method proposed by Zou and Zhang (2009). This simulation method sets the number of parameters (p) which depends on the sample size $(n).$ Let $p = p_n = \left[4n^{1/2}\right] - 5$ for n = 100, 200, 400. The data is generated from the linear regression model

$$\mathbf{y} = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{7}$$

where $\mathbf{y}$ is an $n\times1$ vector of response variable, $\boldsymbol{\beta}$ is a $p\times1$ vector of parameter of regression coefficients, and $\boldsymbol{\varepsilon}$ is an $n\times1$ vector of random errors, $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}),$ where $\sigma = 6.$ Let $\mathbf{X} = \left[\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_p\right]^T$; $\mathbf{X}_j$ is an $n\times1$ vector of the j^{th} predictor variables, $\mathbf{X} \sim N_p(\mathbf{0}, \Sigma),$ where the covariance matrix Σ has the entry $\Sigma_{j,k} = \rho^{|j-k|},\ 1 \le k,\ j \le p.$ We set ρ = 0.5 and 0.8. Let $\mathbf{1}_q$ denotes a $q\times1$ vector of 1's, and $\mathbf{0}_{p-3q}$ denotes a $(p-3q)\times1$ vector of 0's, where $q = [p_n/9].$ Let the true coefficients be $\boldsymbol{\beta} = \left(3\cdot\mathbf{1}_q, 3\cdot\mathbf{1}_q, 3\cdot\mathbf{1}_q, \mathbf{0}_{p-3q}\right)^T.$ Let $A = \{j : \beta_j \ne 0, j = 1,2,...,p\}.$ The size of A is the number of non-zero coefficients which are used to generate the response variable of the model. For this simulation method, the size of A is denoted by $A = 3q.$ There are six cases for combination of n = 100, 200, 400 and ρ = 0.5, 0.8. The simulation method is repeated 1,000 times.

2.2. Comparative study and tuning parameters selection

We study the performance of four different penalized regression methods:

- elastic net method,
- adaptive elastic net method where the adaptive weight is constructed by using rescaled elastic net estimator with $\alpha = 0.9$ as suggested by Jiratchayut (2015), and Jiratchayut and Bumrungsup (2015a) and choosing $\gamma = 3$ as suggested by Zou and Zhang (2009),
- SCAD-L_2 method where the constant $a = 3.7$ (Fan and Li 2001), and
- L1CP method.

These methods consist of the L_1 and L_2 penalties which perform parameter estimation and variable selection. In this research, the elastic net method is fitted using the penalty parameters $(\lambda_1 \text{ and } \lambda_2)$ with $\alpha = 0.5.$ The adaptive elastic net, SCAD-L_2, and L1CP approaches are fitted using the same shrinkage values $(\lambda_1 \text{ and } \lambda_2)$ of the elastic net. The 10-fold cross-validation (CV) method used for tuning the penalty parameters $(\lambda_1 \text{ and } \lambda_2)$ is CV random partition using MATLAB2012a software. The value of λ estimated by 10-fold CV method is the λ with minimum mean prediction squared error as calculated by CV. The elastic net and L1CP methods are implemented using lasso command of MATLAB 2012a software. The adaptive elastic net method is implemented using the gcdnet R package (Yang and Zou 2012, 2015). The SCAD-L_2 method is implemented using the ncvreg R package (Breheny 2017).

2.3. Decision criterion

The elastic net, adaptive elastic net, L1CP, and SCAD-L_2 approaches perform both parameter estimation and variable selection. For this comparative study, the decision criteria are as follows.

1. For each estimator $\hat{\boldsymbol{\beta}},$ its estimation accuracy is measured by the mean square error $(MSE(\hat{\boldsymbol{\beta}}))$ defined as $E\left[\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)^T\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)\right].$

2. The variable selection performance is measured by C and $IC,$ where C is the number of zero coefficients that are correctly estimated by zero and IC is the number of nonzero coefficients that are incorrectly estimated by zero.

3. The prediction accuracy is measured by the prediction error (PE) defined as $E\left(\mathbf{y}-\hat{\mathbf{y}}\right)^2$ where $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}.$

3. Results and Discussion

Tables 1 to 3 show the model selection and fitting results of the naïve elastic net, adaptive elastic net where the adaptive weight is constructed using rescaled elastic net estimator with $\alpha = 0.9,$ L1CP, and SCAD-L_2 for six cases of simulation data. For each method, the average of $PE,$ $MSE(\hat{\boldsymbol{\beta}}),$ $C,$ and IC are computed based on 1,000 datasets. The numbers in parenthesis are the corresponding standard errors of PE and $MSE(\hat{\boldsymbol{\beta}})$ estimated using the bootstrap with $B = 500$ resampling from the $1,000 PE$'s, and $1,000 MSE(\hat{\boldsymbol{\beta}})$'s, respectively.

3.1. Estimation accuracy

For most cases, the adaptive elastic net performs the estimation accuracy better than the other methods. When $n = 100$ and $\rho = 0.8,$ the adaptive elastic net, elastic net, and L1CP perform similarly.

3.2. Variable selection performance

The adaptive elastic net performs the variable selection performance better than the other methods. The L1CP, SCAD-L_2, and elastic net are slightly different in variable selection performance.

3.3. Prediction performance

The SCAD-L_2 performs the prediction performance better than the other methods. When sample size is large, the L1CP performs prediction performance close to the prediction accuracy of the SCAD-L_2.

The elastic net, adaptive elastic net, SCAD-L_2 and L1CP are penalized regression methods which perform both parameter estimation and variable selection. Their penalty functions consist of L_1 and L_2 penalties. The L_1 part is responsible for the sparsity of the estimator by shrinking some coefficients to be zero, while the L_2 part has ability to select group of correlated predictors and stabilizes the solution. The elastic net, adaptive elastic net, and SCAD-L_2 are different in their L_1 part. The adaptive elastic net estimator incorporates the adaptive weight in the L_1 penalty of the elastic net estimator. The SCAD-L_2 is an extension of the elastic net, where the L_1 penalty of the elastic net is replaced by the SCAD function (Fan and Li 2001). Zou and Zhang (2009) studied the performance of the adaptive elastic net and the SCAD (Fan and Li 2001). They found that the adaptive elastic net has parameter estimation and variable selection performance better than the SCAD. In this paper, we compare the SCAD-L_2 (an extension of the SCAD, by adding the L_2 penalty to the SCAD function) (Zeng and Xie 2014) with the elastic net, adaptive elastic net, and L1CP. This simulation study shows that the adaptive elastic net also has parameter estimation and variable selection performance better than the SCAD-L_2. For the L_2 part, the elastic net and L1CP methods are different. The elastic net penalty is a combination of the lasso (L_1) and ridge (L_2) penalties, while the L1CP is a combination between the L_1 norm and a correlation based penalty. The result of this research reveals that the L1CP has prediction performance better than the elastic net, which agrees with Anbari and Mkhadri (2014). For this research, we study the performance of the elastic net, adaptive elastic net, L1CP, and SCAD-L_2 which are different in L_1 and L_2 parts. The result of this research shows that the adaptive elastic net has the parameter estimation and variable selection performance better than the others, especially when p and n are large. This agrees with Zou and Zhang (2009), who claimed that the adaptive elastic net is designed for high-dimensional data analysis. For prediction performance, SCAD-L_2 has the prediction performance better than the others, but SCAD-L_2 and L1CP are slightly different.

Mao and Ye (2017) proposed a penalized linear regression method which is called LqCP. This method performs variable selection and has ability to do group selection. They studied the performance of LqCP, L1CP (Anbari and Mkhadri 2014), elastic net (Zou and Hastie 2005), and adaptive elastic net (Zou and Zhang 2009) where the adaptive weight was constructed by using the elastic net estimator. By considering only elastic net, adaptive elastic net, and L1CP methods, Mao and Ye (2017) showed that the adaptive elastic net has better performance in parameter estimation and variable selection than elastic net and L1CP. Their results are similar to the result of this research. Mao and Ye (2017) did not study the performance of SCAD-L_2.

Table 1 Model selection and fitting results of four different penalized regression estimators for $n = 100$ and $p_n = 35$

ρ	$\lvert A \rvert$	Truth		Model	$MSE(\hat{\boldsymbol{\beta}})$	C	IC	PE
		C	IC					
0.5	9	26	0	Elastic net	0.2575(0.0033)	9.0250	0.0000	28.4693(0.1761)
				Adaptive elastic net	0.2349(0.0034)	18.6220	0.0000	28.7967(0.1562)
				SCAD-L_2	0.3332(0.0047)	9.3960	0.0000	26.7107(0.1663)
				L1CP	0.2555(0.0035)	10.8190	0.0000	27.9837(0.1661)
0.8	9	26	0	Elastic net	0.2439(0.0063)	12.6120	0.0030	29.9723(0.1697)
				Adaptive elastic net	0.2446(0.0065)	20.0550	0.0120	30.0684(0.1664)
				SCAD-L_2	0.3433(0.0085)	13.5620	0.0030	28.5260(0.1684)
				L1CP	0.2388(0.0063)	14.3960	0.0040	29.8000(0.1708)

Table 2 Model selection and fitting results of four different penalized regression estimators for $n = 200$ and $p_n = 51$

ρ	$\lvert A \rvert$	Truth		Model	$MSE(\hat{\boldsymbol{\beta}})$	C	IC	PE
		C	IC					
0.5	15	36	0	Elastic net	0.1482(0.0014)	11.4290	0.0000	29.8578(0.1107)
				Adaptive elastic net	0.1079(0.0012)	29.2660	0.0000	31.2812(0.1069)
				SCAD-L_2	0.2010(0.0019)	11.8650	0.0000	28.5509(0.1055)
				L1CP	0.1586(0.0017)	13.1660	0.0000	28.9864(0.1072)
0.8	15	36	0	Elastic net	0.1318(0.0021)	18.6410	0.0000	31.5066(0.1234)
				Adaptive elastic net	0.1112(0.0022)	31.0070	0.0000	32.2722(0.1123)
				SCAD-L_2	0.1963(0.0033)	19.8750	0.0000	30.2494(0.1278)
				L1CP	0.1604(0.0026)	21.0180	0.0000	30.7772(0.1188)

Table 3 Model selection and fitting results of four different penalized regression estimators for $n = 400$ and $p_n = 75$

ρ	$\lvert A \rvert$	Truth		Model	$MSE(\hat{\boldsymbol{\beta}})$	C	IC	PE
		C	IC					
0.5	24	51	0	Elastic net	0.0828(0.0007)	15.5150	0.0000	31.0127(0.0802)
				Adaptive elastic net	0.0499(0.0005)	45.7860	0.0000	32.9366(0.0813)
				SCAD-L_2	0.1143(0.0009)	16.0060	0.0000	30.0879(0.0825)
				L1CP	0.0907(0.0008)	17.1670	0.0000	30.2146(0.0764)
0.8	24	51	0	Elastic net	0.0782(0.0008)	27.5690	0.0000	32.7482(0.0853)
				Adaptive elastic net	0.0580(0.0007)	47.5180	0.0000	33.8759(0.0802)
				SCAD-L_2	0.1216(0.0015)	29.1970	0.0000	31.6996(0.0873)
				L1CP	0.1105(0.0012)	30.3310	0.0000	31.9431(0.0826)

4. Real Data Examples

We illustrate the comparative study with two real datasets: prostate cancer data (Stamey et al. 1989) and UScrime data (Ehrlich 1973, Vandaele 1978, and Venables and Ripley 1999).

4.1. Prostate cancer data

The prostate cancer data is a data from a prostate cancer studied by Stamey et al. (1989). This data was collected from 97 patients. The response variable is the logarithm of prostate specific antigen (lpsa). The predictor variables are eight clinical measures: the logarithm of cancer volume (lcavol), the logarithm of prostate weight (lweight), age, the logarithm of the amount of benign prostatic hyperplasia (lbph), seminal vesicle invasion (svi), the logarithm of capsular penetration (lcp), the Gleason score (gleason), and the percentage Gleason score 4 or 5 (pgg45). For prostate cancer data, the value $\gamma = 1.7$ is used for computing the adaptive weight of the naïve adaptive elastic net. The naïve elastic net, adaptive elastic net, SCAD-L_2, and L1CP estimators are fitted with the shrinkage parameters $\lambda_1 = 0.0224$ and $\lambda_2 = 0.0224$. Table 4 shows model selection and fitting result of four penalized regression methods. The result reveals that SCAD-L_2 performs the prediction performance better the other methods, and the elastic net has prediction error higher than the SCAD-L_2. This result is similar to the research of Zeng and Xie (2014) which found that SCAD-L_2 has slightly lower prediction error than elastic net.

Table 4 Model selection and fitting result of different penalized regression methods for prostate cancer data

Method	Predictor variables								Degrees of freedom	*PE*
	lcavol	lweight	age	lbph	svi	lcp	gleason	pgg45		
Elastic net	0.4973	0.5699	-0.0117	0.0717	0.6117	0.0000	0.0238	0.0025	7	0.4555
Adaptive elastic net	0.4731	0.5782	0.0000	0.0000	0.6536	0.0000	0.0000	0.0000	3	0.4868
SCAD-L_2	0.5444	0.6091	-0.0193	0.0936	0.7243	-0.0796	0.0000	0.0050	7	0.4451
L1CP	0.5136	0.4830	-0.0012	0.0662	0.6034	0.0000	0.0000	0.0022	6	0.4669

4.2. UScrime data

UScrime data is a data from a study of criminologists who are interested in the effect of punishment regimes on crime rates (Ehrlich 1973, and Vandaele 1978). The UScrime data is available in R package: MASS (Venables and Ripley 1999, and Ripley et al. 2017). This data has been studied using aggregate data on 47 states of the USA in 1960. The response variable is the rate of crimes in a particular category per head of population. There are 15 predictor variables: percentage of males aged 14-24 (M), indicator variable for a Southern state (So), mean years of schooling (Ed), police expenditure in 1960 (Po1), police expenditure in 1959 (Po2), labour force participation rate (LF), number of males per 1,000 females (M.F.), state population (Pop), number of non-whites per 1,000 people (NW), unemployment rate of urban males 14-24 (U1), unemployment rate of urban males 35-39 (U2), gross domestic product per head (GDP), income inequality (Ineq), probability of imprisonment (Prob), and average time served in state prisons (Time).

For UScrime data, the value $\gamma = 5$ is used for computing the adaptive weight of the naïve adaptive elastic net. The naïve elastic net, adaptive elastic net, SCAD-L_2 and L1CP estimators are fitted with the shrinkage parameters $\lambda_1 = 0.2236$ and $\lambda_2 = 0.2236$. Table 5 shows model selection and fitting result of four penalized regression methods. The result reveals that SCAD-L_2 performs the prediction performance better than the other methods. The adaptive elastic net performs prediction performance close to the prediction performance of the SCAD-L_2. From Table 5, the L1CP performs the prediction performance worse than the other methods. It differs from the result of Anbari and Mkhadri (2014) which found that the L1CP is the winner in term of prediction.

5. Conclusions

The elastic net, adaptive elastic net, L1CP, and SCAD-L_2 are different penalized linear regression methods which have the ability to achieve grouping effect. These methods consist of L_1 and L_2 penalties which perform parameter estimation and variable selection. The adaptive elastic net and SCAD-L_2 have the oracle property-consistency in selection and asymptotic normality, but the elastic net and L1CP do not enjoy the oracle property. The results show that the adaptive elastic net performs best in parameter estimation and variable selection performance. For minimizing prediction error, simulation and real datasets reveal that the SCAD-L_2 has the prediction performance better than the others. When sample size is large, the SCAD-L_2 and

L1CP are slightly different in prediction performance. There are two tuning parameters λ_1 and λ_2 in the elastic net, adaptive elastic net, L1CP, and SCAD-L_2 methods. The penalty parameter λ_1 offers the sparsity of the estimator, while the penalty parameter λ_2 is responsible for the solution part and improves the prediction. In this research, the 10-fold cross-validation method is used for tuning the penalty parameters (λ_1 and λ_2). To improve the prediction performance of the adaptive elastic net, the other way for tuning the value of λ_2 is selection the value of λ_2 based on Bayes factor which is proposed by Jiratchayut (2015), and Jiratchayut and Bumrungsup (2015b). Their study showed that the value of λ_2 based on Bayes factor improves the prediction performance of the elastic net and adaptive elastic net (Jiratchayut 2015, and Jiratchayut and Bumrungsup 2015b).

Table 5 Model selection and fitting result of different penalized regression methods for UScrime data

		Method			
		Elastic net	Adaptive elastic net	SCAD-L_2	L1CP
Predictor variables	M	5.2222	5.1756	5.2489	5.3190
	So	89.2750	100.2062	89.8980	0.9880
	Ed	6.7904	7.0229	6.8370	11.7500
	Po1	4.2738	4.6106	4.2710	18.9800
	Po2	3.7982	4.1726	3.7999	-6.4499
	LF	0.6470	0.6962	0.6506	0.9273
	M.F.	2.3197	2.2255	2.3278	-3.8797
	Pop	0.2594	0.0000	0.2624	-2.2656
	NW	0.3645	0.1781	0.3650	-0.2196
	U1	-1.8872	-1.9520	-1.9119	0.0715
	U2	7.5714	8.0542	7.6390	7.1867
	GDP	0.3323	0.0000	0.3373	0.0000
	Ineq	2.4424	2.3331	2.4560	5.1635
	Prob	-3,409.7000	-3,415.2099	-3,415.6086	-0.6920
	Time	1.0059	1.7646	1.0188	0.1322
Degrees of freedom		15	13	15	14
PE		40,103.48	40,021.6	40,004.31	54,367.46

Penalized regression is a shrinkage method for regression analysis where multicollinearity problem exists. This approach is a tool in the fields of data mining and machine learning (Hastie et al. 2009, and James et al. 2013). The elastic net, adaptive elastic net, L1CP, and SCAD-L_2 are penalized linear regression approaches used for linear regression analysis where the predictors show degree of multicollinearity and the regression model encourages a grouping effect, e.g., microarray classification and gene selection. In this paper, we study the performance of the elastic net, adaptive elastic net, L1CP, and SCAD-L_2 in full rank model ($p < n$). For further work, it would be useful to study these methods when $p > n.$ The LqCP (Mao and Ye 2017) is the other penalized regression method which performs variable selection and has ability to do group selection. There is no comparative study between LqCP and SCAD-L_2. This will be a topic for future research.

Acknowledgements

This paper was supported by Faculty of Science and Arts, Burapha University, Chanthaburi Campus. The authors would like to thank the editor and the referees for their helpful comments and suggestions.

References

Anbari ME, Mkhadri A. Penalized regression combining the L_1 norm and a correlation based penalty. Sankhya Ser B. 2014; 76(1): 82-102.

Breheny P. ncvreg: Regularization paths for SCAD and MCP penalized regression models, R package version 3.9-1. 2017 [cited 2017 Aug 10]; 12-14. Available from: https://cran.r-project.org/web/packages/ncvreg/.

Ehrlich I. Participation in illegitimate activities: a theoretical and empirical investigation. J Polit Econ. 1973; 81: 521-565.

Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. J Am Stat Assoc. 2001; 96: 1348-1360.

Ghosh S. On the grouped selection and model complexity of the adaptive elastic net. Stat Comput. 2011; 21(3): 451-462.

Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. New York: Springer; 2009.

Hoerl AE, Kennard RW. Ridge regression: applications to nonorthogonal problems, Technometrics. 1970a; 12: 69-82.

Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. Technometrics. 1970b; 12: 55-67.

James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning with applications in R. New York: Springer; 2013.

Jiratchayut K. Bayesian elastic net regression. PhD [dissertation]. Pathum Thani: Thammasat University; 2015.

Jiratchayut K, Bumrungsup C. A study of adaptive elastic net estimators with different adaptive weights. Sci Tech Asia. 2015a; 20(3): 1-7.

Jiratchayut K, Bumrungsup C. Elastic net regression with the value of the L_2 penalty parameter associated with Bayesian analysis. Thail Stat. 2015b; 13(2): 243-269.

Mao N, Ye W. Group variable selection via a combination of L_q norm and correlation-based penalty. Adv Pure Math. 2017; 7: 51-65.

Ripley B, Venables B, Bates DM, Hornik K, Gebhardt A, Firth D. MASS: support functions and datasets for Venables and Ripley's MASS, R package version 7.3-47. 2017 [cited 2017 Oct 27]; 156-157. Available from: https://cran.r-project.org/.

Stamey TA, Kabalin JN, Mcneal JE, Johnstone IM, Freiha F, Redwine EA, Yang N. Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate: II. Radical prostatectomy treated patients. J Urology. 1989; 141(5): 1076-1083.

Tibshirani R. Regression shrinkage and selection via the lasso. J R Statist Soc B. 1996; 58: 267-288.

Tutz G, Ulbricht J. Penalized regression with correlation based penalty. Stat Comput. 2009; 19(3): 239-253.

Vandaele W. Participation in illegitimate activities: Ehrlich revisited. In: Blumatein A, Cohen J, Nagin D, editors. Deterrence and Incapacitation. US National Academy of Sciences; 1978. p. 270-335.

Venables WN, Ripley BD. Modern applied statistics with S-PLUS. 3rd ed. New York: Springer; 1999.

Yang Y, Zou H. An efficient algorithm for computing the HHSVM and its generalizations. J Comput Graph Stat. 2012; 22(2): 396-415.

Yang Y, Zou H. gcdnet: A generalized coordinate descent (GCD) algorithm for computing the solution path of the hybrid Huberized support vector machine (HHSVM) and its generalization, R package version 1.0.4. 2015 [cited 2017 Aug 4]; 7-10. Available from: https://cran.r-project.org/web/packages/gcdnet/

Zeng L, Xie J. Group variable selection via SCAD-L_2. Statistics. 2014; 48(1): 49-66.

Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Statist Soc B. 2005; 67: 301-320.

Zou H, Zhang HH. On the adaptive elastic-net with a diverging number of parameters. Ann Stat. 2009; 37(4): 1733-1751.