Research Article

A Study of Adaptive Elastic Net Estimators with Different Adaptive Weights

Kanyalin Jiratchayut* and Chinnaphong Bumrungsup
Department of Mathematics and Statistics, Faculty of Science and Technology,
Thammasat University Rangsit Campus,
Khlong Nueng, Khlong Luang, Pathum Thani 12120, Thailand

Abstract

We study the performance of two adaptive elastic net estimation methods where the adaptive weights are constructed using elastic net and least squares estimators. Simulation studies show that two adaptive weights perform differently. When the elastic net estimator is used, the adaptive elastic net performs best in estimation accuracy and variable selection performance. If the least squares estimator is used, the adaptive elastic net has prediction performance better than using the other adaptive weight.

Keywords: adaptive elastic net; elastic net; adaptive weight; L_1-penalty; variable selection.

1. Introduction

The elastic net proposed by Zou and Hastie [1] is a regularization technique to solve the regression problem in microarray genes expression data. The elastic net simultaneously performs automatic variable selection and continuous shrinkage. 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 an $p \times 1$ 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 elastic net is based on a combination of the ridge (L_2) [2,3] and the lasso (L_1) [4] penalties. The elastic net is defined in two stages. The naïve elastic net estimator is first found via

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

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. The final elastic net estimator is taken to be a rescaled version of the naïve estimator,

***Correspondence:** kanyalinn@gmail.com

$$\widehat{\boldsymbol{\beta}}_{\text{elastic net}} = (1+\lambda_2)\widehat{\boldsymbol{\beta}}_{\text{naïve elastic net.}} \quad (3)$$

The scaling was introduced by Zou et al. [1] to reduce perceived overshrinkage of the naïve estimator.

The elastic net has good performance. However, it does not enjoy the oracle properties (consistency in variables selection and asymptotic normality). Zou and Zhang [5], and Ghosh [6] proposed two adaptive elastic net estimators which have the oracle property. These two adaptive elastic net estimators are different.

Zou and Zhang [5] proposed the adaptive elastic net using the elastic net estimator to construct the adaptive weight (AENET2009). This adaptive elastic net has the oracle property and outperforms the elastic net. The naïve adaptive elastic net estimator AENET2009 is defined as follows:

$$\widehat{\boldsymbol{\beta}}_{\text{AENET2009}} = \arg\min_{\boldsymbol{\beta}}\left[\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda_2\|\boldsymbol{\beta}\|_2^2 + \lambda_1\sum_{j=1}^{p}\widehat{w}_j|\beta_j|\right]. \quad (4)$$

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

Ghosh [6] proposed the adaptive elastic net using the least squares estimator to construct the adaptive weight (AENET2011). This method has good performance on grouped selection and model complexity than the elastic net. The naïve adaptive elastic net estimator AENET2011 is defined as follows:

$$\widehat{\boldsymbol{\beta}}_{\text{AENET2011}} = \arg\min_{\boldsymbol{\beta}}\left[\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda_2\sum_{j=1}^{p}\left(|\beta_j| + \frac{\lambda_1}{2\lambda_2}\widehat{w}_j\right)^2 - \frac{(\lambda_1)^2}{4\lambda_2}\sum_{j=1}^{p}\widehat{w}_j^2\right]. \quad (5)$$

For some $\gamma > 0$, the adaptive weight vector $\widehat{\mathbf{w}} = 1/\left|\widehat{\boldsymbol{\beta}}_{least\ square}\right|^{\gamma}$. The adaptive elastic net proposed by Ghosh [6] has good performance on grouped selection and model complexity than the elastic net. There is no any comparative study between two adaptive elastic net methods.

In this research, we study the performance of two adaptive elastic net methods where the adaptive weights are constructed using elastic net and least squares estimators. We limit our attention to full rank model ($p < n$). This article is organized as follows. Section 2 describes the simulation method, adaptive weights used in this comparative study, and decision criterion. Section 3 presents the results and discussion. In Section 4, we illustrate our study using a real dataset. Conclusion is in Section 5.

2. Methods

2.1 Simulation data

The datasets are simulated by the simulation method proposed by Zou and Zhang [5]. This simulation method sets the number of parameters (p) depend 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

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

where $\mathbf{y}$ is an $n \times 1$ vector of response variable, $\boldsymbol{\beta}$ is an $p \times 1$ vector of parameter of regression coefficients, and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of random errors, $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$, where σ = 6.

Let $\mathbf{X} = \left[\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_p\right]^T$; $\mathbf{X}_j$ is an $n \times 1$ vector of the jth predictor variables. $\mathbf{X}$ follows a p-dim multivariate normal distribution with zero mean and covariance $\boldsymbol{\Sigma}$, $\mathbf{X} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$, where the covariance matrix $\boldsymbol{\Sigma}$ has the entry $\boldsymbol{\Sigma}_{j,k} = corr(j,k) = \rho^{|j-k|}$, $1 \le k,\ j \le p$. In this research, we set ρ = 0.5 and ρ = 0.75.

Let $\mathbf{1}_q$ denotes a $q \times 1$ vector of 1's, and $\mathbf{0}_{p-3q}$ denotes a $(p-3q) \times 1$ vector of 0's.

Let the true coefficients are

$$\boldsymbol{\beta} = \left(3 \cdot \mathbf{1}_q, 3 \cdot \mathbf{1}_q, 3 \cdot \mathbf{1}_q, \mathbf{0}_{p-3q}\right)^T$$

and $q = [p_n/9]$. Let $\mathcal{A} = \{j : \beta_j \neq 0, j = 1, 2, \ldots, p\}$. The size of $\mathcal{A}$ 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 $\mathcal{A}$ is denoted by $|\mathcal{A}| = 3q$. There are six cases for combination of n = 100, 200, 400 and ρ = 0.5, 0.75. The simulation method is repeated 100 times.

2.2 Adaptive weight

In this research, there are two types of estimators which are used to construct the adaptive weight: elastic net and least squares estimators. For elastic net estimator, we use rescaled version of the elastic net and the relationship between the shrinkage parameters is $\alpha = \lambda_2/(\lambda_1 + \lambda_2)$ where $\alpha \in (0,1)$, so we set rescaled elastic net estimator with α = 0.1, 0.5, 0.9 to be three estimators. Hence, we study four adaptive weights (w):

- adaptive weight is constructed using ordinary least squares estimator (OLS),
- adaptive weight is constructed using elastic net estimator with α = 0.1 (RENET01),
- adaptive weight is constructed using elastic net estimator with α = 0.5 (RENET05), and
- adaptive weight is constructed using elastic net estimator with α = 0.9 (RENET09).

To construct the adaptive weight, we choose $\gamma > \frac{2\nu}{1-\nu}$ where $\lim_{n\to\infty} \frac{\log p}{\log n} = \nu$ as suggested by Zou and Zhang [5]. Thus, the value $\gamma = 3$ is used for fitting the adaptive elastic net of this simulation data. The naïve adaptive elastic net estimators are fitted using the same shrinkage values (λ_1 and λ_2) of the naïve elastic net method with α = 0.5. The elastic net method is implemented using lasso command of MATLAB2012a software [7,8]. The 10-fold cross-validation (CV) method for tuning the penalty parameters (λ_1 and λ_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 adaptive elastic net method is implemented using the gcdnet R package [9-11]. Both the lasso command of MATLAB2012a software and the gcdnet R package solve the elastic net based on the cyclical coordinate descent algorithms proposed by Friedman, Hastie, and Tibshirani [12].

2.3 Decision criterion

The decision criterions are as follows.

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

2. The variable selection performance is gauged by (C, 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 $E(\mathbf{y} - \hat{\mathbf{y}})^2$ where $\hat{\mathbf{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$.

3. Results and Discussion

Table 1 – Table 3 show the model selection and fitting results of naïve adaptive elastic net estimators with different adaptive weights for six cases of simulation data. The average of $MSE(\widehat{\boldsymbol{\beta}})$, PE, C, and IC are computed based on 100 datasets. The numbers in parenthesis are the corresponding standard errors of $MSE(\widehat{\boldsymbol{\beta}})$ and PE estimated using the bootstrap with B = 500 resampling from the 100 $MSE(\widehat{\boldsymbol{\beta}})$'s, and 100 PE's, respectively. The results are as follows.

3.1 Estimation accuracy

The AENET2009 performs the estimation accuracy better than the AENET2011 and elastic net do.

3.2 Variable selection performance

The AENET2009 performs the variable selection performance better than the elastic net and AENET2011 do.

3.3 Prediction performance

The AENET2011 performs the prediction performance better than the AENET2009 does. For the AENET2009, the adaptive weight RENET09 has the prediction performance better than the other adaptive weights.

The least squares method is the parameter estimation, whereas the classical elastic net method [1] performs both parameter estimation and variable selection. Thus, the adaptive weight derived from the elastic net estimator makes the adaptive elastic net has the parameter estimation and variable selection performance better than the weight derived from least squares estimator. The different shrinkage values of α affect the parameter estimation and variable selection performance of the adaptive elastic net. The AENET2009 with adaptive weight RENET01 has the parameter estimation and variable selection performance better than the other adaptive weights do, but it performs worse in prediction accuracy.

4. Real data example

The data in this example is the diabetes dataset [13]. The response variable is a quantitative measure of disease progression one year after baseline for 442 diabetes patients. The dataset contains 10 baseline predictor variables: age, sex, body mass index (bmi), average blood pressure (bp), and six blood serum measurements (tc, ldl, hdl, tch, ltg, glu). Table 4 shows the result of the naïve elastic net estimators with α = 0.1, 0.5, 0.9. The 10-fold CV method for tuning the penalty parameters (λ_1 and λ_2) is CV random partition. This causes the different values of λ_1 and λ_2 at each value of α. The rescaled elastic net estimators which are used to construct the adaptive weights are computed by (3). For diabetes dataset, the value $\gamma = 1.3$ is used for computing the adaptive weights of the adaptive elastic net estimators.

Table 5 shows the results of the naïve adaptive elastic net estimators with different adaptive weights for diabetes dataset. The naïve adaptive elastic net estimators are fitted using the same shrinkage values (λ_1 and λ_2) of the naïve elastic net method with α = 0.5 (λ_1 = 0.0736, λ_2 = 0.0736). The result reveals that the AENET2011 performs the prediction performance better than the AENET2009 does. For the AENET2009, the adaptive weight RENET09 has the prediction performance better than the other adaptive weights. The AENET2009 is more parsimonious than the elastic net and AENET2011 do.

5. Conclusion

The adaptive elastic net estimators incorporate the adaptive weight in the L_1 penalty of the naïve elastic net estimator. The L_1 penalty is responsible for the sparsity of the estimator. When the elastic net estimator is used to construct the adaptive weight, the adaptive elastic net performs best in estimation accuracy and variable selection performance. If the least squares estimator is used to construct the adaptive weight, the adaptive elastic net has the prediction performance better than using the other adaptive weights. The adaptive elastic net does both parameter estimation and variable selection, so the elastic net estimator is more suitable to construct the adaptive weight than the least squares estimator. Using the elastic net estimator with $\alpha \to 1$ to construct the adaptive weight (e.g. RENET09), the adaptive elastic net has the prediction performance better than using the adaptive weight with $\alpha \to 0$.

Table 1. Model selection and fitting results of naïve adaptive elastic net estimators for n = 100, p_n = 35.

n = 100, p_n = 35							
ρ	Truth		Model	$MSE(\hat{\boldsymbol{\beta}})$	C	IC	PE
	C	IC					
0.5	26	0	Elastic net	0.2649 (0.0113)	8.82	0	28.6017 (0.5310)
			AENET2011 w = OLS	0.3286 (0.0183)	15.24	0.08	27.8393 (0.5022)
			AENET2009 w = RENET01	0.1847 (0.0127)	23.53	0.02	31.4825 (0.5548)
			AENET2009 w = RENET05	0.2115 (0.0125)	21.12	0	30.0953 (0.5515)
			AENET2009 w = RENET09	0.2434 (0.0121)	18.46	0	28.9537 (0.5193)
0.75	26	0	Elastic net	0.2239 (0.0158)	11.98	0	29.8884 (0.5297)
			AENET2011 w = OLS	0.4387 (0.0287)	13.71	0.28	28.0819 (0.4469)
			AENET2009 w = RENET01	0.1778 (0.0180)	24.50	0.09	32.1040 (0.4735)
			AENET2009 w = RENET05	0.1891 (0.0183)	22.45	0.02	30.9953 (0.5119)
			AENET2009 w = RENET09	0.2191 (0.0165)	19.95	0	29.9943 (0.4850)

Table 2. Model selection and fitting results of naïve adaptive elastic net estimators for n = 200, p_n = 51.

n = 200, p_n = 51							
ρ	Truth		Model	$MSE(\hat{\boldsymbol{\beta}})$	C	IC	PE
	C	IC					
0.5	36	0	Elastic net	0.1451 (0.0042)	11.68	0	29.4533 (0.3756)
			AENET2011 w = OLS	0.1272 (0.0043)	26.70	0	30.0657 (0.3534)
			AENET2009 w = RENET01	0.0823 (0.0031)	34.39	0	32.4601 (0.3442)
			AENET2009 w = RENET05	0.0940 (0.0036)	31.90	0	31.5969 (0.3397)
			AENET2009 w = RENET09	0.1069 (0.0039)	29.47	0	30.8325 (0.3555)
0.75	36	0	Elastic net	0.1310 (0.0058)	17.33	0	31.4804 (0.3823)
			AENET2011 w = OLS	0.1802 (0.0099)	24.05	0.05	30.8438 (0.3566)
			AENET2009 w = RENET01	0.0866 (0.0051)	35.07	0	33.7111 (0.3689)
			AENET2009 w = RENET05	0.0929 (0.0056)	33.33	0	33.1468 (0.3616)
			AENET2009 w = RENET09	0.1071 (0.0057)	30.82	0	32.4236 (0.3521)

Table 3. Model selection and fitting results of naïve adaptive elastic net estimators for n = 400, p_n = 75.

n = 400, p_n = 75							
ρ	Truth		Model	$MSE(\hat{\boldsymbol{\beta}})$	C	IC	PE
	C	IC					
0.5	51	0	Elastic net	0.0811 (0.0021)	15.57	0	30.6452 (0.2618)
			AENET2011 w = OLS	0.0536 (0.0016)	44.15	0	32.2205 (0.2482)
			AENET2009 w = RENET01	0.0412 (0.0012)	49.99	0	33.4250 (0.2473)
			AENET2009 w = RENET05	0.0447 (0.0014)	48.16	0	32.9821 (0.2616)
			AENET2009 w = RENET09	0.0490 (0.0015)	46.04	0	32.5572 (0.2693)
0.75	51	0	Elastic net	0.0807 (0.0022)	24.68	0	32.3491 (0.2344)
			AENET2011 w = OLS	0.0782 (0.0027)	41.00	0	32.6870 (0.2333)
			AENET2009 w = RENET01	0.0512 (0.0017)	50.48	0	34.4035 (0.2207)
			AENET2009 w = RENET05	0.0530 (0.0018)	49.33	0	34.1404 (0.2268)
			AENET2009 w = RENET09	0.0572 (0.0018)	47.01	0	33.7126 (0.2264)

Table 4. Naïve elastic net estimators for diabetes dataset with α = 0.1, 0.5, 0.9.

The naïve elastic net estimators ($\hat{\beta}$) with different α values													
α	λ_2	λ_1	Predictor variables										Degree of freedom	PE
			AGE	BMI	BP	S1	S2	S3	S4	S5	S6	SEX		
0.9	0.0537	0.0059	-0.0080	5.4561	1.0692	-0.1798	-0.0654	-0.6519	4.1757	43.6550	0.3303	-20.9670	10	2880.171
0.5	0.0736	0.0736	0	5.3747	1.0506	-0.1382	-0.0907	-0.6865	4.0241	41.9670	0.3391	-20.1910	9	2885.256
0.1	0.0633	0.5699	0	5.3863	1.0217	-0.1136	-0.0648	-0.7369	2.5814	42.2340	0.3026	-18.7210	9	2890.408

Table 5. Naïve adaptive elastic net estimators with different adaptive weights for diabetes dataset.

Model	Predictor variables										Degree of freedom	PE
	AGE	BMI	BP	S1	S2	S3	S4	S5	S6	SEX		
Elastic net α = 0.5	0	5.3747	1.0506	-0.1382	-0.0907	-0.6865	4.0241	41.9670	0.3391	-20.1910	9	2885.256
AENET2011 w = OLS	0	5.4247	1.0721	-0.0695	-0.1616	-0.7442	4.7600	36.9364	0.3702	-20.5848	9	2892.505
AENET2009 w = RENET09	0	5.4230	0.9997	0	-0.2038	-0.8348	3.6166	38.4832	0.3577	-20.0363	8	2892.913
AENET2009 w = RENET05	0	5.4379	0.9784	0	-0.2038	-0.8333	3.5934	38.6564	0.3620	-19.9281	8	2893.479
AENET2009 w = RENET01	0	5.4578	0.9560	0	-0.2044	-0.8295	3.6287	38.6405	0.3685	-19.8173	8	2894.476

6. Acknowledgements

The authors gratefully acknowledge the financial support provided by Thammasat University Research Fund under the TU Research Scholar, Contract No.30/2557.

7. References

[1] Zou, H., and Hastie, T., Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society, Series B*, Vol.67, pp. 301-320, 2005.

[2] Hoerl, A. E., and Kennard, R. W., Ridge Regression: Applications to Nonorthogonal Problems, *Technometrics*, Vol.12, No.1, pp.69-82, 1970.

[3] Hoerl, A. E., and Kennard, R. W., Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics* (b), Vol.12, No.1, pp. 55-67, 1970.

[4] Tibshirani, R., Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society, Series B*, Vol.58, pp. 267-288, 1996.

[5] Zou, H., and Zhang, H. H., On the Adaptive Elastic-net with a Diverging Number of Parameters. *The Annals of Statistics*, Vol.37, pp. 1733-1751, 2009.

[6] Ghosh, S., On the Grouped Selection and Model Complexity of the Adaptive Elastic Net. *Statistics and Computing*, Vol.21, pp. 451-462, 2011.

[7] MATLAB and Simulink, Lasso and Elastic Net, Available from URL: http://www.mathworks.com/help/stats/lasso-and-elastic-net.html (accessed February 13, 2013).

[8] MATLAB and Simulink, Regularized Least-squares Regression Using Lasso or Elastic Net Algorithm, Available from URL:http://www.mathworks.com/help/stats/lasso.html (accessed February 13, 2013).

[9] Yang, Y., and Zou, H., An Efficient Algorithm for Computing the HHSVM and Its Generalizations, *Journal of Computational and Graphical Statistics*, Vol.22, pp. 396-415, 2012.

[10] Yang, Y., and 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 1.0.3, 2013, URL: http://code.google.com/p/gcdnet/

[11] R Development Core Team, R: *A language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2013, URL: http://www.r-project.org/

[12] Friedman, J., Hastie, T., and Tibshirani, R., Regularization Paths for Generalized linear Models via Coordinate Descent. *Journal of Statistical Software*, Vol.33, No.1, pp. 1-22, 2010.

[13] Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R., Least Angle Regression, *The Annals of Statistics*, Vol.32, pp. 407-499, 2004.