

Article

The Higher-Order of Adaptive Lasso and Elastic Net Methods for Classification on High Dimensional Data

Autcha Araveeporn 

Department of Statistics, King Mongkut's Institute of Technology Ladkrabang, School of Science, Bangkok 10520, Thailand; autcha.ar@kmitl.ac.th

Abstract: The lasso and elastic net methods are the popular technique for parameter estimation and variable selection. Moreover, the adaptive lasso and elastic net methods use the adaptive weights on the penalty function based on the lasso and elastic net estimates. The adaptive weight is related to the power order of the estimator. Normally, these methods focus to estimate parameters in terms of linear regression models that are based on the dependent variable and independent variable as a continuous scale. In this paper, we compare the lasso and elastic net methods and the higher-order of the adaptive lasso and adaptive elastic net methods for classification on high dimensional data. The classification is used to classify the categorical data for dependent variable dependent on the independent variables, which is called the logistic regression model. The categorical data are considered a binary variable, and the independent variables are used as the continuous variable. The high dimensional data are represented when the number of independent variables is higher than the sample sizes. For this research, the simulation of the logistic regression is considered as the binary dependent variable and 20, 30, 40, and 50 as the independent variables when the sample sizes are less than the number of the independent variables. The independent variables are generated from normal distribution on several variances, and the dependent variables are obtained from the probability of logit function and transforming it to predict the binary data. For application in real data, we express the classification of the type of leukemia as the dependent variables and the subset of gene expression as the independent variables. The criterion of these methods is to compare by the average percentage of predicted accuracy value. The results are found that the higher-order of adaptive lasso method is satisfied with large dispersion, but the higher-order of adaptive elastic net method outperforms on small dispersion.



Citation: Araveeporn, A. The Higher-Order of Adaptive Lasso and Elastic Net Methods for Classification on High Dimensional Data. *Mathematics* **2021**, *9*, 1091. <https://doi.org/10.3390/math9101091>

Academic Editor: Panagiotis-Christos Vassiliou

Received: 21 April 2021

Accepted: 10 May 2021

Published: 12 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: elastic net; high dimensional data; lasso

1. Introduction

The regression analysis is a statistical method for the estimation of the relationship between a dependent variable and one or more independent variables. The model from regression analysis is used to predict a continuous dependent variable from several independent variables. If the dependent variable is of a dichotomous scale, then the logistic regression should be used to explain the relationship between a binary dependent variable and one or more independent variables.

The use of logistic regression analysis is focused to predict whether or not an event occurred such as failure or success, diseased or healthy, yes or no. Especially, decision making of the patient is widely used to study in health science for classifying this event. The application of the logistic regression model obtained a cohort of the pregnant woman and the factor that influences the decision to opt for caesarean delivery or vaginal birth [1]. The logistic regression analysis is used to evaluate the effect of the number of events per variable from patients in which deaths occurred [2].

However, parameter estimation is one process for predicting the probability of logit function, which then classifies the categorical classes of the dependent variable. There

are several methods to approximate the parameters such as the maximum likelihood method [3], ridge regression [4], and the Markov Chain Monte Carlo method [5]. These methods estimate parameter for classification of the dependent variable dependent on all independent variables, but they fail to select the important independent variable. When the high dimensional data occur, the selection variable methods are proposed to solve this problem.

The lasso (least absolute shrinkage and selection operator) method [6] is a new method for estimation in the linear model, and it shrinks some coefficients to zero. It tries to retain the good features of subset selection. After that, Zou and Hastie [7] proposed the elastic net that encouraged a grouping effect, in which strongly correlated independent variables tend to be in or out of the model together. The simulation study has shown that the elastic net outperforms the lasso. These methods are the variable selection methods for estimating parameters dependent on the penalty function and tuning parameter.

The developed method is concerned with the adaptive weights and the power of the adaptive weight on the penalty function. Zou [8] proposed the adaptive lasso, where the adaptive weights were used for penalizing different coefficients in the lasso penalty. The adaptive lasso enjoyed the oracle properties by utilizing the adaptive weights and leads to an optimal estimator on the generalized linear model. Zou and Zhang [9] considered the model selection and estimation in high dimensional data called the adaptive elastic net. This method combined the strengths of the quadratic regularization and the adaptive weights lasso shrinkage.

The point of the previous methods is to estimate parameters on the regression model for predicting the continuous dependent variables. To explain the categorical dependent variable, the logistic regression model is used to model the probability of a binary class. Furthermore, we apply the parameter estimation of the lasso, elastic net, and adaptive lasso and elastic methods to solve these problems. The new study is to concentrate on the higher-order of the adaptive lasso and elastic net to estimate parameters and predict the binary dependent variable when the data are in form of high dimensional data. The gene expression is applied to these methods for classification and gene selection. Algamal and Lee [10] studied cancer classification and gene selection on high dimensional data by adjusting adaptive elastic net. They were successful to tackle both estimating the gene coefficients and performing gene selection simultaneously. Zou and Hastie [11] proposed the penalized logistic regression for cancer classification; this method had the advantage of providing an estimate of the underlying probability. Furthermore, there is no comparison of the performance of adaptive lasso and adaptive elastic net methods, which play on the higher-order of adaptive weight in logistic regression model under high dimensional data and apply for a classification of leukemia patient.

The propose of this study is to compare the lasso, elastic net, adaptive lasso, and adaptive elastic net to shrinkage the number of the independent variables and classify the binary dependent variable under the high dimensional data. The higher-order of adaptive lasso and adaptive elastic net is appropriate also to mention the estimated parameter. Moreover, we compare the performance of four methods in logistic regression with the average percentage of predicted accuracy, and the average number of selected variables. These outcomes are obtained to shrinkage the independent variables via a Monte Carlo simulation.

2. Methods

2.1. The Logistic Regression Model

The general class of logistic regression is written by

$$y_i = \pi(x_i) + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (1)$$

where y_i denotes the value of a dichotomous outcome variable, $\pi(x_i)$ denotes the probability of the Bernoulli distribution depended on independent variable, x_i , and ε_i is

called the error and follows a normal distribution with mean zero and variance equal to $\pi(x_i) \left[1 - \pi(x_i) \right]$.

The specific form of the logistic regression model is considered as the probability by

$$\pi(x_i) = \frac{e^{x_i \beta}}{1 + e^{x_i \beta}} = \frac{1}{1 + e^{-x_i \beta}}. \quad (2)$$

The transformation of $\pi(x_i)$ is a central of logistic regression model that is called logit function. This transformation is defined as

$$g(x_i) = \log \frac{\pi(x_i)}{1 - \pi(x_i)} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} = x_i \beta, \quad (3)$$

where β is the vector of coefficient in terms of logit transformation on $(k+1) \times 1$, x_i is a matrix of independent variable on $n \times (k+1)$, k is a number of independent variable, and $i = 1, 2, \dots, n$ is the number of observations. If y_i is coded as 0 or 1, then the probability $\pi(x_i)$ provides $y_i = 1$. It follows that the probability $1 - \pi(x_i)$ gives $y_i = 0$. The probability distribution function to contribute the likelihood function is expressed as

$$P(Y_i = y_i) = \left(\pi(x_i) \right)^{y_i} \left(1 - \pi(x_i) \right)^{1-y_i}, \quad y_i = 0, 1. \quad (4)$$

The likelihood function is obtained from the terms of (4) as

$$l(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i}. \quad (5)$$

The likelihood from (5) can be expressed by taking log as

$$\begin{aligned} L(\beta) &= \ln l(\beta) = \ln \left[\prod_{i=1}^n \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i} \right] \\ &= \sum_{i=1}^n \left[y_i \ln(\pi(x_i)) + (1 - y_i) \ln(1 - \pi(x_i)) \right] \end{aligned} \quad (6)$$

The log-likelihood function in Equation (6) can be written in the form of penalized function as follows [10]:

$$L^*(\beta) = -L(\beta) + \lambda J(\beta), \quad (7)$$

where λ is the tuning parameter and $J(\beta)$ is the penalty function.

2.2. LASSO Method

The least absolute shrinkage and selector operator (Lasso) [6] is a method for estimation parameters in the linear model by minimizing the residual sum of the square to the sum of the absolute values of the coefficients. The shrinkage of the lasso composes by assigning some coefficients to zero and hence tries to retain the good features for variable selection. The lasso estimate β is define by

$$\hat{\beta}_{j(L)} = \underset{\beta}{\operatorname{argmin}} \left[\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij})^2 + \lambda \sum_{j=1}^k |\beta_j| \right]$$

where $\lambda \sum_{j=1}^k |\beta_j|$ is the penalty function.

For the binary dependent variable, the lasso estimate β is regularized from (6) and (7) as:

$$\hat{\beta}_{j(L)} = \underset{\beta}{\operatorname{argmin}} \left[- \sum_{i=1}^n \left\{ y_i \ln(\pi(x_i)) + (1 - y_i) \ln(1 - \pi(x_i)) \right\} + \lambda \sum_{j=1}^k |\beta_j| \right]. \quad (8)$$

The tuning parameter λ is used to try out different values by the cross-validation method. Lasso estimators for all values of λ can be computed through a modification of the least-angle regression (LARS) algorithm [12] which is an algorithm for fitting linear regression model to high dimensional data.

Zou [8] proposed the adaptive lasso, where adaptive weights are used for penalizing different coefficients in the penalty function. The adaptive lasso estimators are defined as

$$\hat{\beta}_{j(AL)} = \underset{\beta}{\operatorname{argmin}} \left[\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij})^2 + \lambda \sum_{j=1}^k \hat{w}_j |\beta_j| \right],$$

where the $\lambda \sum_{j=1}^k \hat{w}_j |\beta_j|$ is the penalty function.

The adaptive lasso can be applied to classify a dichotomous outcome variable, thus, the adaptive lasso estimate β is regularized from (6) and (7) as:

$$\hat{\beta}_{j(AL)} = \underset{\beta}{\operatorname{argmin}} \left[- \sum_{i=1}^n \left\{ y_i \ln(\pi(x_i)) + (1 - y_i) \ln(1 - \pi(x_i)) \right\} + \lambda \sum_{j=1}^k \hat{w}_j |\beta_j| \right]. \quad (9)$$

The adaptive weight is $\hat{w}_j = \frac{1}{|\hat{\beta}_{j(L)}|^\gamma}$; $j = 1, 2, \dots, k$, $\gamma > 0$, where $\hat{\beta}_{j(L)}$ is obtained from (8). The positive value γ is the power of the adaptive weight, which is concerned in higher-order. The tuning parameter λ and the order of adaptive weight γ are used as the two-dimensional cross-validation to tune the adaptive lasso.

2.3. Elastic Net Method

Zou and Hastie [7] proposed the elastic net as a new regularization and variable selection method. The elastic net of penalty function had the characteristics of both lasso and ridge regression [4]. In the parameter estimating, the elastic net estimator found the ridge regression coefficient and made the lasso shrinkage along with the lasso coefficient solution. The elastic net estimate β is define by

$$\hat{\beta}_{j(E)} = \underset{\beta}{\operatorname{argmin}} \left[\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij})^2 + \lambda_1 \sum_{j=1}^k |\beta_j| + \lambda_2 \sum_{j=1}^k \beta_j^2 \right];$$

$$0 < \lambda_1 + \lambda_2 < 1,$$

where $\lambda_1 \sum_{j=1}^k |\beta_j| + \lambda_2 \sum_{j=1}^k \beta_j^2$ is the penalty function.

To consider the classification of y_i , the elastic net estimate β is regularized from (6) and (7) as:

$$\hat{\beta}_{j(E)} = \underset{\beta}{\operatorname{argmin}} \left[- \sum_{i=1}^n \left\{ y_i \ln(\pi(x_i)) + (1 - y_i) \ln(1 - \pi(x_i)) \right\} + \lambda_1 \sum_{j=1}^k |\beta_j| + \lambda_2 \sum_{j=1}^k \beta_j^2 \right]. \quad (10)$$

From (10), the elastic net estimation is a ridge regression when λ_1 is zero as $\hat{\beta}_R = \underset{\beta}{\operatorname{argmin}} \left[-\sum_{i=1}^n \left\{ y_i \ln(\pi(x_i)) + (1 - y_i) \ln(1 - \pi(x_i)) \right\} + \lambda_2 \sum_{j=1}^k \beta_j^2 \right]$. The lasso estimator is in form of (8) when λ_2 is zero. The tuning parameters of λ_1 and λ_2 control the shrinkage of $\hat{\beta}^E$ using cross-validation [13].

The adaptive elastic net is developed from the elastic net to solve the problem when the number of parameters diverges with the sample sizes and the dimension is high. Zou and Zhang [6] proposed the adaptive elastic net that combines the strengths of the quadratic regularization and adaptively weighted lasso shrinkage. The adaptive elastic net is defined as follows:

$$\hat{\beta}_{j(AE)} = \underset{\beta}{\operatorname{argmin}} \left[\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij})^2 + \lambda_1 \sum_{j=1}^k \hat{w}_j |\beta_j| + \lambda_2 \sum_{j=1}^k \beta_j^2 \right].$$

The adaptive elastic net reduces to the adaptive lasso when λ_2 closes to be zero. The penalty function combines the elastic net and adaptive lasso method, and then, the tuning parameter is checked using Bayesian information criterion (BIC) cross-validation [14] which is the method to select an optimal value of regularization parameter.

If we focus to classify a dichotomous outcome variable, then the adaptive elastic net of β can be constructed from (6) and (7) as:

$$\hat{\beta}_{j(AE)} = \underset{\beta}{\operatorname{argmin}} \left[-\sum_{i=1}^n \left\{ y_i \ln(\pi(x_i)) + (1 - y_i) \ln(1 - \pi(x_i)) \right\} + \lambda_1 \sum_{j=1}^k \hat{w}_j |\beta_j| + \lambda_2 \sum_{j=1}^k \beta_j^2 \right]. \quad (11)$$

The adaptive weight is $\hat{w}_j = \left(|\hat{\beta}_{j(E)}| + \frac{1}{n} \right)^{-\gamma}; j = 1, 2, \dots, k, \gamma > 0$, where $\hat{\beta}_{j(E)}$ is obtained from (10). The γ is power of the adaptive weight, which is concentrated as the adaptive lasso.

3. Simulation Data and Results

In the following examples, we compare the classification methods consisted of the lasso, adaptive lasso, elastic net, and adaptive elastic net. The adaptive lasso and elastic net use the higher-order on the adaptive weights. For simulation, we generate the data in form of high dimension data when the number of independent variables is higher than the sample sizes (n). The independent variables test sets by 20 ($n = 15$), 30 ($n = 15, 20, 25$), 40 ($n = 20, 25, 30, 35$), and 50 ($n = 20, 25, 30, 35, 40$) variables from the normal distribution with mean (μ) and variance (σ^2) denoted by $N(\mu, \sigma^2)$. The variance is studied in sets of 1, 5, and 10 that are presented in Figure 1.

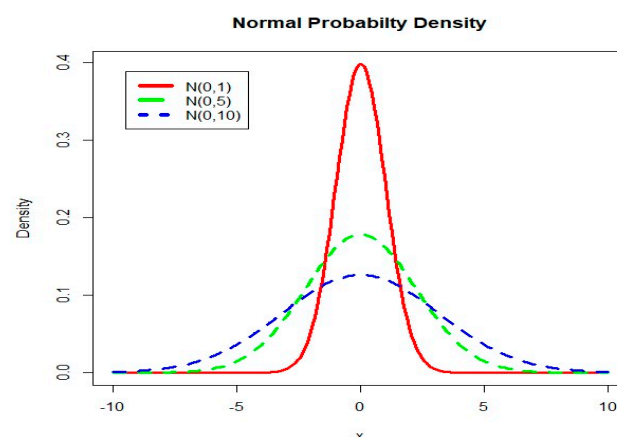


Figure 1. The normal probability density in $\mu = 0$ and variance $\sigma^2 = 1, 5$, and 10.

The coefficients of logistic regression are set the constant values as β , and the classification of dependent variable (y_i) related the probability form as $\pi(x_i) = \frac{1}{1+e^{-x_i\beta}}$. The y_i is noted as 0 or 1, then the probability $\pi(x_i) \geq 0.5$ provides $y_i = 1$, and the probability $\pi(x_i) < 0.5$ gives $y_i = 0$. After that, we obtain the estimator from the previous section as $\hat{\beta}$, and the probability can approximate by $\hat{\pi}(x_i) = \frac{1}{1+e^{-x_i\hat{\beta}}}$. The categorical dependent variables are predicted by $\hat{y}_i = 1$ as $\hat{\pi}(x_i) \geq 0.5$ and by $\hat{y}_i = 0$ as $\hat{\pi}(x_i) < 0.5$. This process can be seen in Figure 2.

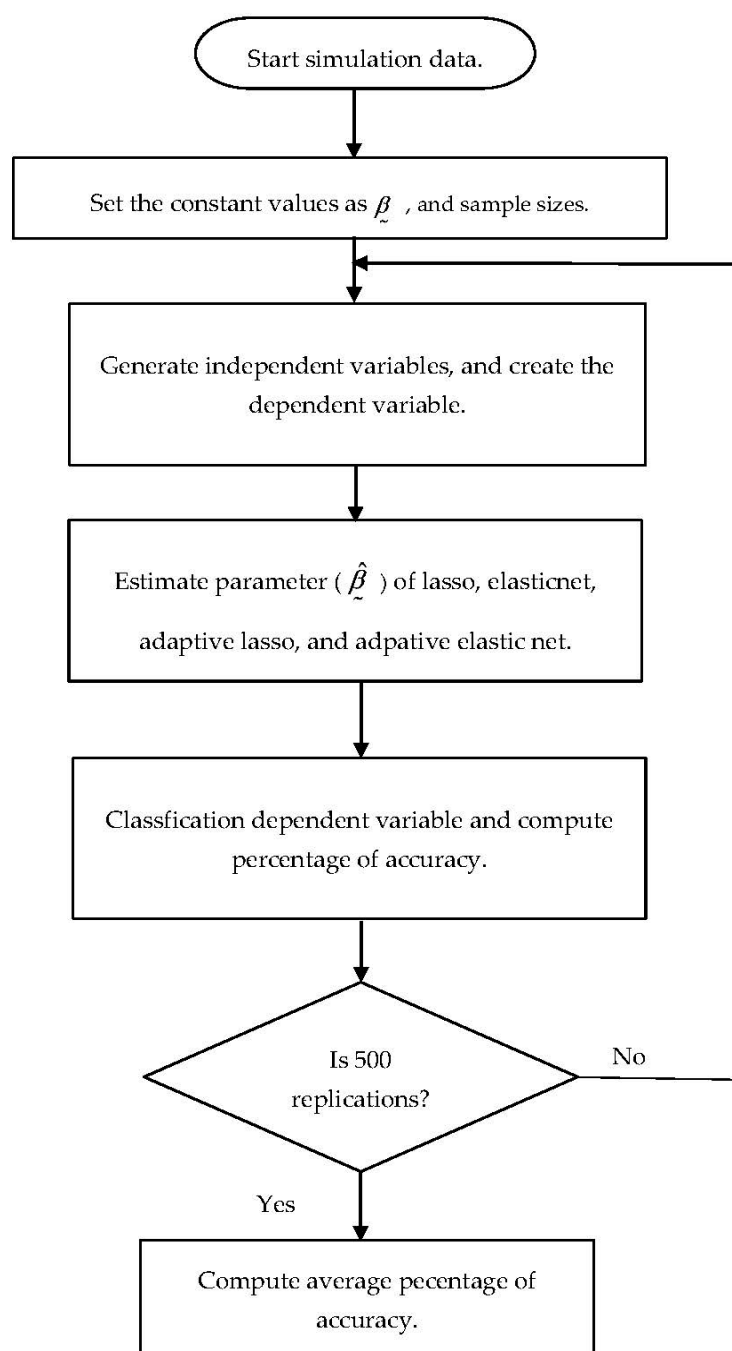


Figure 2. The flowchart diagram of simulation process.

The performance of predictive analytics is the confusion matrix, which compared predicted values and actual data in Table 1.

Table 1. The confusion matrix of actual data (y_i) and predicted data (\hat{y}_i).

Predicted Data	Actual Data	
	$y_i=1$	$y_i=0$
$\hat{y}_{ii} = 1$	True positive (TP)	False positive (FP)
$\hat{y}_i = 0$	False negative (FN)	True negative (TN)

The confusion matrix is used when there are two or more classes as the output of the classifier. The predicted accuracy is computed from Table 1 using

$$\text{Percentage of Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100.$$

The average percentage of accuracy and number of selected variables of the lasso, adaptive lasso, elastic net, and adaptive elastic net under 20, 30, 40, and 50 variables is presented in Tables 2–5. The average percentage of accuracy and number of selected variables is computed by the mean of over 500 replications. For the adaptive lasso and elastic net, the higher-order is denoted by γ . For this process, we use the R program for simulation data and the R package called glmnet and HDeconometrics, which support estimating parameters on these methods. The code of commands presents in the Appendix A.

Table 2. The average percentage of accuracy and the average number of selected variables (#) of lasso, adaptive lasso (AL), elastic net, and adaptive elastic net (AEN) under 20 variables.

Sample Sizes (n)	Normal $N(\mu, \sigma^2)$	Lasso	AL $\gamma = 1$	AL $\gamma = 2$	Elastic Net	AEN $\gamma = 1$	AEN $\gamma = 2$
15	N(0,1) (#)	87.11 (3)	96.20 (4)	99.18 (4)	87.48 (6)	95.71 (5)	99.59 (6)
	N(0,5) (#)	87.14 (5)	96.85 (4)	99.53 (5)	88.17 (10)	91.30 (8)	94.09 (8)
	N(0,10) (#)	87.60 (5)	96.97 (5)	99.46 (5)	88.16 (8)	90.08 (7)	91.40 (8)

Table 3. The average percentage of accuracy and the average number of selected variables (#) of lasso, adaptive lasso (AL), elastic net, and adaptive elastic net (AEN) under 30 variables.

Sample Sizes (n)	Normal	Lasso	AL $\gamma = 1$	AL $\gamma = 2$	Elastic Net	AEN $\gamma = 1$	AEN $\gamma = 2$
15	N(0,1) (#)	84.15 (4)	95.94 (5)	99.54 (5)	85.45 (8)	96.61 (7)	99.87 (7)
	N(0,5) (#)	85.06 (4)	95.78 (5)	99.65 (5)	86.18 (9)	90.10 (8)	94.13 (7)
	N(0,10) (#)	85.02 (4)	95.80 (5)	99.61 (9)	86.76 (8)	88.36 (8)	90.86 (8)
20	N(0,1) (#)	87.90 (6)	97.52 (7)	99.69 (7)	88.32 (10)	97.12 (10)	99.95 (10)
	N(0,5) (#)	87.84 (6)	97.46 (7)	99.63 (7)	88.46 (10)	91.62 (9)	94.83 (9)
	N(0,10) (#)	87.81 (6)	97.32 (7)	99.69 (7)	88.42 (10)	90.14 (9)	91.60 (9)
25	N(0,1) (#)	88.99 (8)	97.75 (9)	99.04 (9)	90.00 (12)	96.63 (12)	99.83 (12)
	N(0,5) (#)	88.76 (9)	97.71 (9)	99.52 (9)	90.85 (12)	93.44 (12)	96.00 (11)
	N(0,10) (#)	88.80 (8)	97.79 (9)	99.53 (9)	90.88 (12)	92.16 (12)	93.60 (12)

Table 4. The average percentage of accuracy and the average number of selected variables (#) of lasso, adaptive lasso (AL), elastic net, and adaptive elastic net (AEN) under 40 variables.

Sample Sizes (n)	Normal	Lasso	AL $\gamma = 1$	AL $\gamma = 2$	Elastic Net	AEN $\gamma = 1$	AEN $\gamma = 2$
20	N(0,1) (#)	87.00 (6)	96.79 (6)	99.49 (7)	97.86 (10)	97.76 (10)	99.97 (10)
	N(0,5) (#)	86.13 (6)	97.39 (6)	99.63 (7)	97.82 (10)	92.01 (10)	95.80 (9)
	N(0,10) (#)	86.23 (5)	97.44 (6)	99.71 (7)	88.14 (10)	90.27 (10)	92.19 (9)
25	N(0,1) (#)	86.73 (7)	97.80 (9)	99.94 (9)	87.14 (12)	97.10 (12)	99.97 (12)
	N(0,5) (#)	86.57 (7)	97.68 (9)	99.70 (9)	87.50 (12)	90.88 (11)	95.75 (11)
	N(0,10) (#)	86.27 (7)	97.72 (9)	99.70 (9)	87.69 (12)	89.04 (12)	92.48 (11)
30	N(0,1) (#)	89.08 (9)	98.18 (11)	99.70 (11)	90.32 (13)	97.57 (13)	99.92 (14)
	N(0,5) (#)	88.19 (9)	98.32 (11)	99.62 (11)	89.76 (14)	93.14 (14)	96.25 (13)
	N(0,10) (#)	88.51 (9)	98.06 (11)	99.72 (11)	89.72 (14)	91.72 (14)	93.50 (13)
35	N(0,1) (#)	89.20 (10)	98.18 (13)	99.70 (13)	90.32 (15)	97.57 (15)	99.92 (17)
	N(0,5) (#)	88.26 (10)	98.32 (12)	99.62 (13)	89.84 (14)	93.14 (15)	96.25 (14)
	N(0,10) (#)	88.59 (10)	98.06 (12)	99.72 (13)	89.78 (14)	91.72 (14)	93.50 (14)

Table 5. The average percentage of accuracy and the average number of selected variables (#) of lasso, adaptive lasso (AL), elastic net, and adaptive elastic net (AEN) under 50 variables.

Sample Sizes (n)	Normal	Lasso	AL $\gamma = 1$	AL $\gamma = 2$	Elastic Net	AEN $\gamma = 1$	AEN $\gamma = 2$
20	N(0,1) (#)	85.18 (6)	96.42 (6)	99.70 (7)	86.09 (11)	98.37 (10)	99.97 (10)
	N(0,5) (#)	85.12 (5)	96.57 (6)	99.72 (7)	86.14 (10)	91.35 (9)	95.82 (9)
	N(0,10) (#)	85.26 (5)	96.60 (6)	99.71 (7)	86.57 (10)	88.14 (10)	99.50 (10)
25	N(0,1) (#)	85.70 (7)	97.73 (8)	99.91 (9)	86.76 (12)	98.00 (12)	99.99 (12)
	N(0,5) (#)	85.29 (7)	97.88 (8)	99.72 (9)	86.64 (12)	91.70 (11)	96.04 (11)
	N(0,10) (#)	84.96 (7)	97.84 (8)	99.82 (9)	86.69 (12)	89.47 (11)	92.63 (11)
30	N(0,1) (#)	84.85 (8)	98.32 (10)	99.77 (11)	86.82 (14)	98.03 (14)	99.98 (14)
	N(0,5) (#)	85.53 (9)	98.18 (11)	99.82 (11)	87.42 (15)	91.54 (13)	95.84 (13)
	N(0,10) (#)	85.73 (9)	97.98 (11)	99.81 (11)	86.96 (14)	89.47 (13)	92.11 (13)
35	N(0,1) (#)	86.40 (10)	98.18 (12)	99.76 (13)	87.30 (16)	97.67 (16)	99.99 (17)
	N(0,5) (#)	86.21 (11)	98.25 (13)	100 (13)	87.37 (16)	92.43 (16)	96.09 (15)
	N(0,10) (#)	85.99 (10)	98.28 (12)	100 (13)	87.52 (16)	90.50 (16)	93.10 (15)
40	N(0,1) (#)	89.22 (12)	98.53 (15)	99.85 (15)	90.85 (18)	97.95 (18)	99.95 (19)
	N(0,5) (#)	88.47 (12)	98.39 (15)	99.50 (15)	89.85 (19)	93.38 (17)	96.26 (17)
	N(0,10) (#)	87.75 (12)	98.37 (15)	99.49 (15)	89.41 (19)	91.84 (17)	93.60 (17)

From Tables 2–5, it can be seen that the adaptive elastic net showed a maximum average percentage of accuracy at the small dispersion when the order is defined at 2. When the dispersion is large, i.e., the adaptive lasso is of good performance depended on the high order, then the average percentage of accuracy of this method is closely related to high variance. An increase in the average percentage of accuracy causes an increase in the number of independent variables.

To concentrate on the average number of selected variables, when the variance is increasing value, there is no effects with selected variables. For the increase in sample sizes, the average number of selected variables is also increasing too. Furthermore, increasing order of adaptive lasso and the elastic net is slightly different number of selected variables.

4. Application in Real Data

To evaluate our proposed methods, gene expression monitoring (via DNA microarray) is used to classify 72 patients with acute myeloid leukemia and acute lymphoblastic leukemia. These data have consisted of 3571 genes from the bone marrow samples, which is described in detail by Golup et al. [15]. A subset of 3571 genes is the independent

variables and simulated by setting the independent variables larger than the sample sizes (n) in 500 replications. The set of independent variables is 20 ($n = 15$), 30 ($n = 15, 20, 25$), 40 ($n = 20, 25, 30, 35$), and 50 ($n = 20, 25, 30, 35, 40$), which is similar the simulation data. The sample sizes are selected from 72 patients under simple random sampling. The example of leukemia genes is shown in Figure 3, which can be seen in the descriptive statistics on the box plot diagram. The multicollinearity of genes is presented in Figure 4 in terms of correlation matrix among 40 sample genes. The average percentage of accuracy and the average number of selected genes of the lasso, adaptive lasso, elastic net, and adaptive elastic net under 20, 30, 40, and 50 leukemia genes can be found in Tables 6–10.

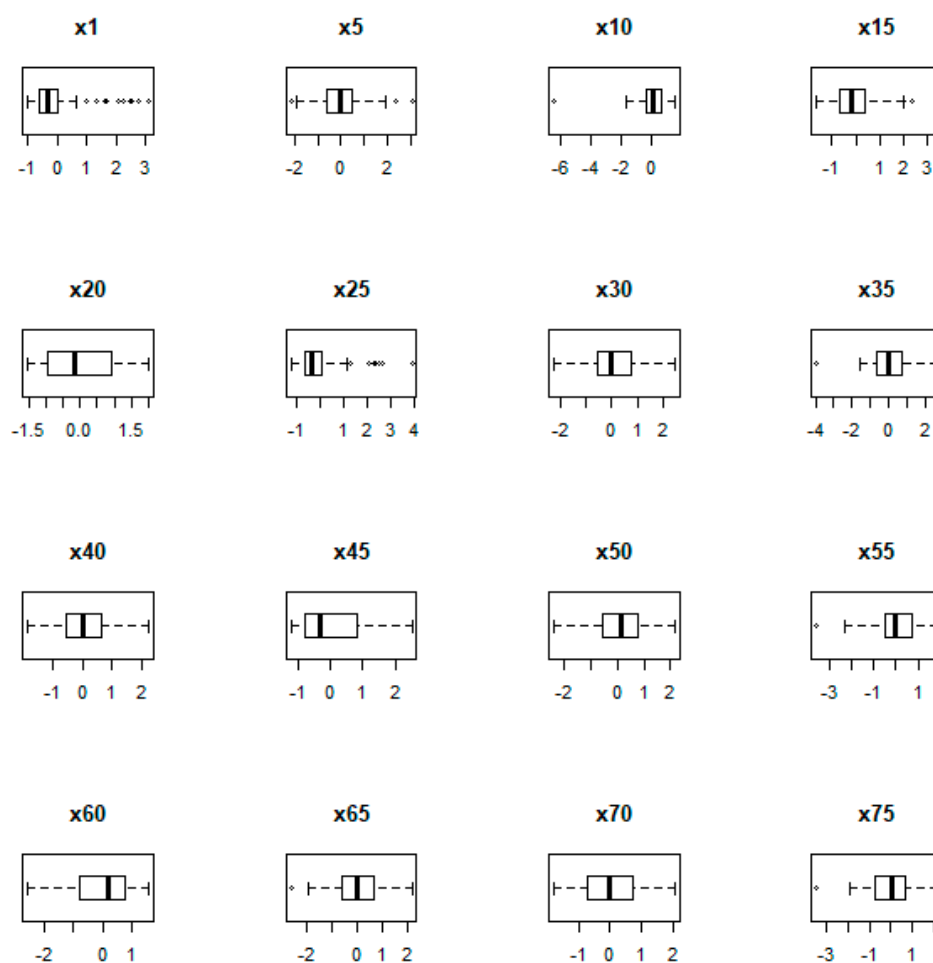


Figure 3. The boxplot of sample leukemia genes.

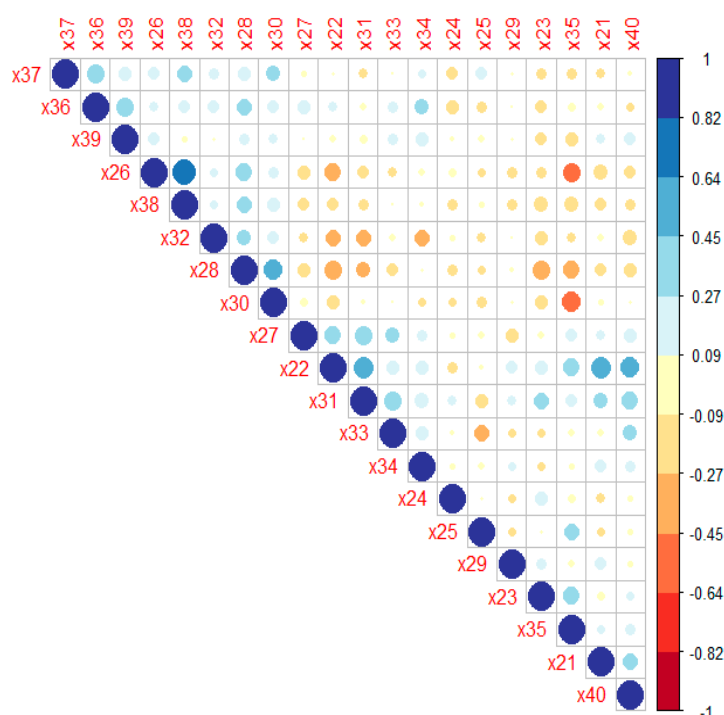


Figure 4. The graph of correlation matrix of sample leukemia genes.

Table 6. The average percentage of accuracy and the average number of selected genes of lasso, adaptive lasso (AL), elastic net, and adaptive elastic net (AEN) under 20 leukemia genes.

Sample Sizes (<i>n</i>)	Lasso	AL $\gamma = 1$	AL $\gamma = 2$	Elastic Net	AEN $\gamma = 1$	AEN $\gamma = 2$
15	93.95	94.01	96.39	95.10	95.21	97.05
# Selected genes	(9)	(4)	(4)	(12)	(9)	(8)

Table 7. The average percentage of accuracy and the average number of selected genes of lasso, adaptive lasso (AL), elastic net, and adaptive elastic net (AEN) under 20 leukemia genes.

Sample Sizes (<i>n</i>)	Lasso	AL $\gamma = 1$	AL $\gamma = 2$	Elastic Net	AEN $\gamma = 1$	AEN $\gamma = 2$
15	93.95	94.01	96.39	95.10	95.21	97.05
# Selected genes	(9)	(4)	(4)	(12)	(9)	(8)

Table 8. The average percentage of accuracy and the average number of selected genes of lasso, adaptive lasso (AL), elastic net, and adaptive elastic net (AEN) under 30 leukemia genes.

Sample Sizes (<i>n</i>)	Lasso	AL $\gamma = 1$	AL $\gamma = 2$	Elastic Net	AEN $\gamma = 1$	AEN $\gamma = 2$
15	96.12	95.57	97.96	97.40	97.42	98.39
20	95.68	95.73	98.06	97.59	97.70	98.62
25	95.41	96.03	98.35	98.10	98.41	98.54
# Selected genes	(12)	(5)	(5)	(15)	(11)	(10)

Table 9. The average percentage of accuracy and the average number of selected genes of lasso, adaptive lasso (AL), elastic net, and adaptive elastic net (AEN) under 40 leukemia genes.

Sample Sizes (<i>n</i>)	Lasso	AL $\gamma = 1$	AL $\gamma = 2$	Elastic Net	AEN $\gamma = 1$	AEN $\gamma = 2$
20	96.47	96.68	98.70	97.99	99.06	99.03
25	96.08	96.21	98.68	98.01	98.24	98.94
30	95.41	96.03	98.35	98.01	98.41	98.54
35	91.99	93.83	97.65	97.17	97.66	97.78
# Selected genes	(13)	(5)	(5)	(17)	(11)	(10)

Table 10. The average percentage of accuracy and the average number of selected genes of lasso, adaptive lasso (AL), elastic net, and adaptive elastic net (AEN) under 50 leukemia genes.

Sample Sizes (<i>n</i>)	Lasso	AL $\gamma = 1$	AL $\gamma = 2$	Elastic Net	AEN $\gamma = 1$	AEN $\gamma = 2$
20	96.51	96.68	99.08	98.64	98.64	99.40
25	96.66	97.02	99.22	98.80	98.90	99.47
30	96.26	95.56	98.92	98.66	98.85	99.15
35	93.24	94.77	98.22	98.06	98.21	98.59
40	93.85	95.20	98.47	98.47	98.59	98.65
# Selected genes	(14)	(5)	(6)	(20)	(11)	(11)

In Figure 3, the bold line in the box plot presents the middle value of the dataset, called median, and the whiskers indicate variability outside the first and the third quartiles. The dot lines drawn horizontally from the box represented the minimum and maximum of all data. Outliers may be plotted as individual points in x1, x5, x10, x15, x25, x35, x55, x65, and x75. It can be seen that the median is closed to zero and the variance is one.

The correlation matrix shows the different shades. The light shade presents a weak correlation, and the dark shade plays a strong correlation. From Figure 4, it can be seen that the weak correlation following the light shade is larger than the dark shade. It means that the sample dataset has a slight correlation. However, all genes are selected for computing parameter, so it reduces the multicollinearity problem.

As can be seen from Tables 6–10, the higher-order of adaptive elastic net achieves a maximum average percentage of accuracy for 50 leukemia genes. Furthermore, it is clear from the results that the adaptive elastic net outperforms the lasso and adaptive lasso in terms of classification accuracy for these data sets. To mention the selection variables, the elastic net selects more genes than the other methods for all data sets, but the adaptive lasso selects fewer genes than the other methods. The adaptive has the potential to select genes that made the highest accuracy. The classification of the type of leukemia patient can use some gene expressions that it can save the time and budget to collect large data. The study of gene selection and classification has proposed the other techniques to detect the expression level of thousands of genes in a few experiments [16–18].

5. Discussion

From the simulated results in Tables 2–5, the factors influencing the average percentage of accuracy are the variance level of the independent variable, sample sizes, and the order of adaptive weights. An increase in the variance affects a decrease in the average percentage of accuracy in most cases. In the case of sample sizes, if the sample size is increased then the accuracy of all methods decreases in all cases. Moreover, an increase in the high order causes an increase in the average percentage of accuracy for all variables.

For the results of gene expression monitoring in Tables 6–10, the sample sizes and the order of adaptive weights are affected similarly to the simulation data. On the other hand, the adaptive elastic net shows the highest average percentage of accuracy. It is obvious that DNA microarray real data demonstrate the small variance as seen from Figure 1. Again,

we compute the mean of 3571 genes and present in the histogram in Figure 5, which is collected from the 72 patients. Then, t -test is used to confirm that the mean of all genes is equal to zero significance. From hypothesis testing, the p -value (0.0466) is higher than the significant level (0.01), then we accept the null hypothesis $H_0 : \mu = 0$. It can be concluded that the mean of the gene expression is equal to zero. The mean and variance are similar to the simulation dataset as mean equal to zero and variance equal to one. Overall, it is clear that the adaptive elastic net is a good performance for classification of the type of leukemia for some genes.

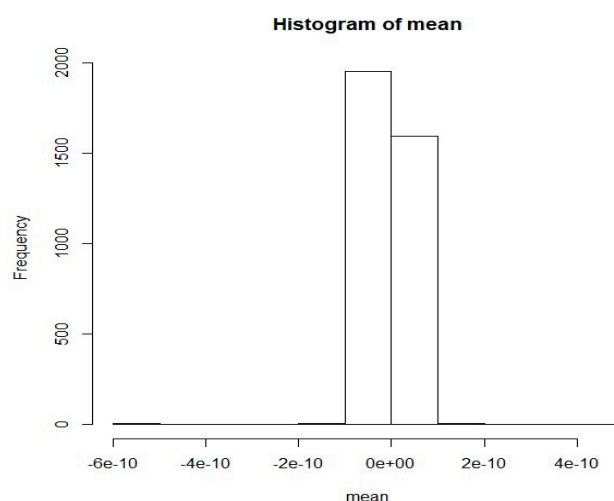


Figure 5. The mean of 3571 leukemia genes presented by histogram.

6. Conclusions

We have proposed the lasso, elastic net, adaptive lasso, and adaptive elastic methods for estimating parameters for classification in binary data. For the empirical results on the high dimensional data, the adaptive elastic net of higher-order is performed in classification when the small dispersion is used in case of simulation data, but adaptive lasso of higher-order is provided in the large dispersion. When using the selection variable, these methods tend to reduce the large independent variables to outperform. For application to actual data, the gene expression data are used to classify the type of leukemia patients. The high dimensional data are concerned in this case because the number of patients is less than the large gene expression data. The simulation of thousands of gene expression are used to compute the percentage accuracy, and these methods are also used to select the influential variables. From the results, it was shown that the adaptive elastic net is effective in gene selection, which was based on the dispersion of data and higher-order of adaptive weights. Therefore, we can conclude that the higher-order is beneficial to classification.

Funding: This research was funded by King Mongkut's Institute of Technology Ladkrabang.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This research is supported by King Mongkut's Institute of Technology Ladkrabang.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

The following commands were used to fit the logistic regression model,
 ##### x is the independent variables #####
 #####

```
##### y is the dependent variable#####
#####
#####The Lasso Method#####
#####
library(glmnet)
cvfit = cv.glmnet(x,y, alpha = 1)
ll = cvfit$lambda.min
fit_las = glmnet(x,y, family = "binomial", lambda = ll,alpha = 1)
est_las = predict(fit_las, newx = x)
#####The Adaptive Lasso#####
#####
library(HDeconometrics)
tau = 1
lasso=ic.glmnet(x,y,crit = "bic")
first.step.coef=coef(lasso)[-1]
penalty.factor=abs(first.step.coef)^(-tau)
cvfit = cv.glmnet(x,y, alpha = 1)
ll = cvfit$lambda.min
fit_Alass1 = glmnet(x,y, family = "binomial", alpha = 1, lambda = ll, penalty.factor =
penalty.factor)
est_Alass1 = predict(fit_Alass1, newx = x)
#####Elastic Net#####
#####3
cvfit = cv.glmnet(x,y, alpha = 0.5)
ll = cvfit$lambda.min
fit_elas = glmnet(x,y, family = "binomial", lambda = ll,alpha = 0.5)
est_elas = predict(fit_elas, newx = x)
#####The Adaptive Elastic Net#####
#####
library(HDeconometrics)
tau = 1
lasso = ic.glmnet(x,y,crit = "bic")
first.step.coef=coef(lasso)[-1]
penalty.factor=(abs(first.step.coef)+(1/(nrow(x))))^(-tau)
cvfit = cv.glmnet(x,y, alpha = 0.5)
ll = cvfit$lambda.min
fit_Aelastic = glmnet(x,y, family = "binomial", alpha = 0.5, lambda = ll, penalty.factor
= penalty.factor)
est_Aelas = predict(fit_Aelastic, newx = x)
#####
#####
```

References

1. Boateng, E.Y.; Abaye, D.A. A review of the logistic regression model with emphasis on medical research. *J. Data Anal. Inf. Process.* **2019**, *7*, 190–207. [\[CrossRef\]](#)
2. Peduzzi, P.; Concato, J.; Kemper, E.; Holford, T.R.; Feinstein, A.R. A simulation study of the number of events per variable in logistic regression analysis. *J. Clin. Epidemiol.* **1996**, *49*, 1373–1379. [\[CrossRef\]](#)
3. Duffy, D.E.; Santner, T.J. On a small sample properties of norm-restricted maximum likelihood estimators for logistic regression models. *Commun. Stat. Theory Methods* **1989**, *18*, 959–980. [\[CrossRef\]](#)
4. Hoerl, A.E.; Kennard, R.W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **1970**, *12*, 55–67. [\[CrossRef\]](#)
5. Araveeporn, A.; Klomwiset, Y. The estimated parameter of logistic regression model by Markov Chain Monte Carlo method with multicollinearity. *Stat. J. IAOS* **2020**, *36*, 1253–1259. [\[CrossRef\]](#)
6. Tishirani, R. Regression shrinkage and selection via lasso. *J. R. Stat. Soc. Ser. B* **1996**, *58*, 267–288.
7. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B* **2005**, *67*, 301–320. [\[CrossRef\]](#)

8. Zou, H. The adaptive lasso and its oracle properties. *J. Am. Stat. Assoc.* **2006**, *101*, 1418–1429. [[CrossRef](#)]
9. Zou, H.; Zhang, T. On the adaptive elastic net with a diverging number of parameters. *Ann. Stat.* **2009**, *37*, 1733–1751. [[CrossRef](#)] [[PubMed](#)]
10. Algamal, Z.Y.; Lee, M.H. Regularized logistic regression with adjusted adaptive elastic net for gene selection in high dimensional cancer classification. *Comput. Biol. Med.* **2015**, *67*, 136–145. [[CrossRef](#)] [[PubMed](#)]
11. Zou, J.; Hastie, T. Classification of gene microarrays by penalized logistic regression. *Biostatistics* **2004**, *3*, 427–443.
12. Efron, B.; Hastie, T.; Johnson, I.; Tibshirani, R. Least angle regression. *Ann. Stat.* **2004**, *32*, 407–499. [[CrossRef](#)]
13. Hastie, T.; Tibshirani, T.; Friedman, J.B. *The Elements of Statistical Learning: Data Mining Inference and Prediction*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2009; pp. 241–247.
14. Zou, H.; Hastie, T.; Tibshirani, R. On the Degrees of Freedom of Lasso. *Ann. Stat.* **2007**, *35*, 2173–2192. [[CrossRef](#)]
15. Golub, T.R.; Solnín, D.K.; Tamayo, P.; Huard, C.; Gaaswinkbeek, M.; Mesirov, J.P.; Coller, K.; Loh, M.L.; Downing, J.R.; Caligiuri, M.A.; et al. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science* **1999**, *286*, 532–537. [[CrossRef](#)] [[PubMed](#)]
16. Kastrin, A.; Peterlin, B. Rash-Based High-Dimensionality Data Reduction and Class Prediction with Applications to Microarray Gene Expression Data. *Expert Syst. Appl.* **2010**, *37*, 5178–5185. [[CrossRef](#)]
17. Chandra, B.; Gupta, M. An Efficient Statistical Feature Selection Approach for Classification of Gene Expression Data. *J. Biomed. Inform.* **2011**, *44*, 529–535. [[CrossRef](#)] [[PubMed](#)]
18. Lotfi, E.; Keshavarz, A. Gene Expression Microarray Classification using PCA-BEL. *Comput. Biol. Med.* **2014**, *54*, 180–187. [[CrossRef](#)] [[PubMed](#)]