P-ISSN 2586-9000
E-ISSN 2586-9027
Homepage : *https://tci-thaijo.org/index.php/SciTechAsia* Science & Technology Asia
Vol.28 No.2 April - June 2023 Page: [31-42]

Original research article

Comparisons of Penalized Regression Methods under High-Dimensional Sparse Data with Correlated Variables

Supranee Lisawadi, Apisara Sripanich, Patchanok Srisuradetchai*

Department of Mathematics and Statistic, Faculty of Science and Technology, Thammasat University, Pathum Thani 12120, Thailand

Received 6 January 2023; Received in revised form 10 March 2023
Accepted 22 March 2023; Available online 14 June 2023

ABSTRACT

Regression models are frequently used to explain a response variable using independent variables in statistics. However, it is common to encounter situations in which the number of independent variables exceeds the number of observations and predictors are correlated. In this instance, a large number of predictors are statistically insignificant, also known as sparse data. Standard statistical methods do not always apply to such data. Problematic aspects include interpretation, estimation inefficiency, and computation. The penalized regression method, which consists of Ridge, least absolute shrinkage and selection operator (LASSO), elastic net (Enet), adaptive LASSO (ALASSO), and adaptive elastic net (AEnet), is frequently employed during the estimation and variable selection phases. The purpose of this paper was to assess the prediction and variable selection performances of Ridge, Enet, LASSO, ALASSO, and AEnet methods in multiple linear regression with normal or positively skewed error terms, sparse data, and correlated independent variables. In addition, Poisson and logistic regression models are studied. The adaptive weights are created using the remaining three estimators: Ridge, Enet, and LASSO. The results indicate that the Ridge estimator is a viable initial adaptive weight estimator for ALASSO and AEnet. In terms of prediction, AEnet and ALASSO typically outperform the competition. Given the objectives, different tactics are necessary to achieve the lowest false positive rate (FPR) and false negative rate (FNR). Enet or AEnet is essential to attain the lowest FPR, while LASSO or ALASSO will yield the lowest FNR.

Keywords: Adaptive elastic net; Adaptive LASSO; Adaptive weights; Penalized regression

***Corresponding author:** patchanok@mathstat.sci.tu.ac.th **doi:** 10.14456/scitechasia.2023.24

1. Introduction

A generalized linear model (GLM) is a statistical model that can be used to analyze a wide variety of data types, such as count, categorical, and binary data, which is not limited to whether the data must be continuous or normal distributed. GLM enables the fitting of regression models for univariate response data that follow a very general distribution known as the exponential family, which has the general form:

$$f(y|\theta,\phi)=\exp\left[\left(y\theta-b(\theta)\right)/a(\phi)+c(y,\phi)\right],$$

where θ is called the canonical parameter and represents the location, while the dispersion parameter, ϕ, represents the scale and defines various members of the family by specifying the function [1]. Currently, the expansion of data is in two dimensions: the number of independent variables (p) and the number of observations (n). When p exceeds n, as in bioinformatics, genetics, or geography, statistical analysis becomes more challenging. Many traditional statistical methods are ineffective with these types of data, and high-dimensional data can result in highly correlated independent variables or a problem known as multicollinearity [2]. Maximum likelihood (ML) and ordinary least squares (OLS) estimations, for example, become unreliable or impossible to compute directly due to high dimensionality and collinearity problems [3].

Data reduction is one method for dealing with high-dimensional data and correlated independent variables. As a result, selecting variables is critical for detecting significant predictors. In such cases, penalized estimation methods, such as shrinking estimators towards the zeroes vector [3] or performing variable selection and coefficient estimation simultaneously, are often beneficial, as they sacrifice a little bias to reduce the variance of the predicted values and may improve overall prediction accuracy, resulting in an easily interpretable model [4]. The estimated coefficients are derived by minimizing the objective function:

$$\hat{\boldsymbol{\beta}}=\arg\min_{\boldsymbol{\beta}}\left\{\sum_{i=1}^{n}\left(y_i-\sum_{j=1}^{p}\beta_j x_j\right)^2+P_\lambda(\boldsymbol{\beta})\right\},$$

or $\hat{\boldsymbol{\beta}}=\arg\min_{\boldsymbol{\beta}}\left\{-l(\boldsymbol{\beta})+P_\lambda(\boldsymbol{\beta})\right\}$,

where $l(\boldsymbol{\beta})$ is a loglikelihood function and $P_\lambda(\boldsymbol{\beta})$ is the penalty function.

Hoerl and Kennard [5] proposed a new multi-collinearity method, namely, Ridge regression. Ridge regression reduces coefficients and is more stable than OLS. It does not set coefficients to zero or produce a clear model. Tibshirani [4] proposed the least absolute shrinkage operator estimator (LASSO). Some coefficients are shrunk, and others are reset. LASSO is popular for high-dimensional data, but it has some flaws [2]. When the number of independent variables is much larger than the number of observations $(p >> n)$ and a grouped variables situation, LASSO is not the ideal method, because it can only select at most n variables out of p candidates. To overcome these limitations, Zou and Hastie [6] proposed the elastic net (Enet), which is particularly useful when $p >> n$. Following that, Zou [7] introduced an adaptive LASSO (ALASSO), in which adaptive weights are utilized to penalize distinct coefficients in the penalty function and enjoy the oracle property -- consistency in variable selection and asymptotic normality [8].

In addition, Zou and Zhang [9] proposed the adaptive elastic net (AEnet) that combines the elastic net and the ALASSO, which both have the oracle property. From the foregoing, it can be noted that the ALASSO and AEnet regression methods have adaptive weight in the penalty function. Many previous studies have constructed the adaptive weights differently. For example, Choosawat, Reangsephet *et al.* [10], Sarakor and Kulvanich [11], Lisawadi *et al.* [12], Neammai and Lisawadi [13] used Ridge estimator, Rungsaranon and Araveeporn [14] used LASSO and elastic net estimator, and Porndumnernsawat and Hirunkasi [15] used $X_j'Y/n$ to construct the adaptive weights. Jiratchayut and Bumrungsup [8] studied the performance of two AEnet estimation methods where the adaptive weights are constructed using elastic net and OLS estimators, and they found that the two adaptive weights perform differently.

Consequently, the purpose of this paper is to compare the efficacy of penalized regression estimators such as Ridge, LASSO, elastic net (Enet), adaptive LASSO (ALASSO), and adaptive elastic net (AEnet) regression methods in many situations. The adaptive weights required for ALASSO and AEnet are determined using three estimators: the Ridge, LASSO, and Enet estimators. The GLMs and multiple linear regression models are of interest for high-dimensional sparse data with correlated independent variables. The constant, Toeplitz, and Hub Toeplitz correlation structures are used to determine correlation among many independent variables. These situations have not been studied in the literature we reviewed.

2. Materials and Methods

There are three components in GLMs: (i) random component which is the distribution of dependent variable $Y_i, i = 1,2,\ldots,n$: a member of an exponential family, such as normal, binomial, Poisson, gamma, or inverse-Gaussian distributions, (ii) systematic component which is a linear function of regressors $\boldsymbol{\eta} = \mathbf{X}^{\mathbf{T}}\boldsymbol{\beta} = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}$, and (iii) link function, $g(\cdot)$, which transforms the expectation of the dependent variable, $\mu_i = E(Y_i)$, to the linear predictors: $g(\mu_i) = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$.

Table 1. Some common link functions for the GLM.

Distribution	Link function
normal	$\eta_i = \mu_i$ (identity link)
binomial	$\eta_i = \ln\left(\pi_i / (1-\pi_i)\right)$ (logistic link)
Poisson	$\eta_i = \ln(\mu_i)$ (log link)
exponential	$\eta_i = 1/\mu_i$ (reciprocal link)
gamma	$\eta_i = 1/\mu_i$ (reciprocal link)

The maximum likelihood (ML) method is the theoretical foundation for parameter estimation in GLMs. In high-dimensional data, however, ML estimation becomes inaccurate or difficult to directly compute, and when the independent variables are highly correlated, it leads to instability and excessive variance.

2.1 Estimation strategies

In this paper, we examine the penalized regression method (PR), which simultaneously estimates regression coefficients and selects variables. The PR is always employed with sparse, independent, and high-dimensional data, and it can connect the penalty function, $\lambda P(\boldsymbol{\beta})$, to the residual sum of squares (RSS) or log-likelihood function to provide a more accurate estimation of the prediction error by preventing overfitting [2]. The objective function, $g(\boldsymbol{\beta})$ of PR, is defined as Eq. (2.1), which is equivalent to Eq. (2.2):

$$g(\boldsymbol{\beta}) = \sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p}\beta_j x_j\right)^2 + \lambda P(\boldsymbol{\beta}), \qquad (2.1)$$

$$g(\boldsymbol{\beta}) = -l(\boldsymbol{\beta}) + \lambda P(\boldsymbol{\beta}), \qquad (2.2)$$

where $l(\boldsymbol{\beta})$ is a loglikelihood function and λ is a tuning parameter, which controls the strength of the penalty and shrinks the coefficients $\boldsymbol{\beta}$ towards zero vector [16]. The estimators of the vector $\boldsymbol{\beta}$ are obtained by minimizing Eq. (2.1) or (2.2), and the results are the following:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} \beta_j X_j \right)^2 + \lambda P(\boldsymbol{\beta}) \right\}$$

or

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left\{ -l(\boldsymbol{\beta}) + \lambda P(\boldsymbol{\beta}) \right\}. \tag{2.3}$$

It reduces to the OLS and ML estimators when $\lambda = 0$, respectively.

2.1.1 Ridge Estimator

Hoerl and Kennard [5] proposed Ridge regression as a new technique for dealing with multicollinearity. Ridge regression is biased but has a lower variance than the OLS estimator. The Ridge penalty function is given as follows:

$$P_\lambda(\boldsymbol{\beta}) = \lambda \sum_{j=1}^{p} \beta_j^2, \tag{2.4}$$

where $\lambda \geq 0$ is the tuning parameter. In practice, the value of λ is determined by using 5-fold cross-validation.

2.1.2 LASSO

The least absolute shrinkage and selection operator estimator (LASSO) was first proposed by Tibshirani [4]. It shrinks some coefficients and sets others to zero and hence tries to retain the good features of both subset selection and Ridge regression [4]. Conversely, LASSO shrinkage produces biased estimates for large coefficients and thus could be suboptimal in terms of estimation risk [9]. The penalty function of LASSO is defined as follows:

$$P_\lambda(\boldsymbol{\beta}) = \lambda \sum_{j=1}^{p} \left| \beta_j \right|, \tag{2.5}$$

where $\lambda \geq 0$ is a tuning parameter.

2.1.3 Elastic net Estimator

When the number of predictors exceeds the number of observations, an Enet is extremely useful. It simultaneously performs automatic variable selection and continuous shrinkage and can select groups of correlated variables [6]. The penalty function of an elastic net is defined as follows:

$$P_\lambda(\boldsymbol{\beta}) = \lambda_1 \sum_{j=1}^{p} \left| \beta_j \right| + \lambda_2 \sum_{j=1}^{p} \beta_j^2, \tag{2.6}$$

where λ_1 and λ_2 are tuning parameters. As a result, the elastic net method includes the LASSO and Ridge regression, i.e., each of them is a special case where $\lambda_1 = \lambda, \lambda_2 = 0$ or $\lambda_1 = 0, \lambda_2 = \lambda$ [3].

2.1.4 Adaptive LASSO Estimator

Zou [7] proposed adaptive LASSO (ALASSO), where adaptive weights are used for penalizing different coefficients in the penalty. The adaptive LASSO enjoys oracle properties; namely, it performs as well as if the true underlying model were given in advance. The basic idea behind ALASSO is assigning a higher weight to the small coefficients and lower weight to the larger coefficients [17]. This does not imply that LASSO will perform better in forecasting than ALASSO, but the latter will be superior in variable selection. The penalty function of ALASSO is defined as follows:

$$P_\lambda(\boldsymbol{\beta}) = \lambda \sum_{j=1}^{p} \hat{w}_j \left| \beta_j \right|, \tag{2.7}$$

where $\lambda \geq 0$ is a tuning parameter and $\hat{w}_j = \left| \hat{\beta}_j \right|^{-\gamma}$ is the adaptive weight based on the initial estimator $\hat{\beta}_j$, where γ is a positive constant and is usually set to equal one and $\hat{\beta}_j$ is an initial estimator.

2.1.5 Adaptive elastic net Estimator

Adaptive elastic net (AEnet) was first proposed by Zou and Zhang [9]. It can be viewed as a combination of the Enet and ALASSO. The penalty functionmof ALASSO is defined as follows:

$$P_\lambda(\boldsymbol{\beta}) = \lambda_1 \sum_{j=1}^{p} \hat{w}_j |\beta_j| + \lambda_2 \sum_{j=1}^{p} \beta_j^2, \qquad (2.8)$$

where $\lambda_j \geq 0, j = 1,2,...,p$ is the tuning parameter and $\hat{w}_j = |\hat{\beta}_j|^{-\gamma}$ is the adaptive weight based on the initial estimator.

2.2 Simulation study

To conduct ALASSO and AEnet regression methods, adaptive weights are first needed. Weights can be constructed using three estimators: Ridge, LASSO, and Enet. There is no weight for the ridge and Enet methods. Then, the efficiency of regression coefficient estimation and variable selection by penalized regression were evaluated using Monte Carlo simulations with 1,000 replications.

A distribution of independent variables does not affect parameter estimation as these variables are not random variables. Without loss of generality, predictors are generated by the multivariate normal distribution with zero mean and covariance $\sum_{p\times p}$. We consider $\sum_{p\times p}$ with three different correlation structures, namely, constant, Toeplitz, and Hub Toeplitz, respectively. They are as follows:

$$\sum_{j,k} = \rho, 1 \leq k, j \leq p, \qquad (2.9)$$

$$\sum_{j,k} = \rho^{|j-k|}, \qquad (2.10)$$

and $\alpha_{k,1} = 1, \alpha_{k,i} = \rho_{\max} - \tau_k (i-2)$, (2.11)

where $\tau_k = (\rho_{\max} - \rho_{\min})/(p-2)$ and ρ is pairwise correlation.

Let the number of independent variables (p) be $200, 500, 1000$, and the sample size (n) be 50 and 100. The pairwise correlation is set to $\rho = 0.5$ and $\rho = 0.9$. The true parameter vector is specified as $\boldsymbol{\beta} = (1, -0.5, -0.5, 0.1, 0.1, 0.1, \overbrace{0.05, \ldots, 0.05}^{4}, \underbrace{0.01, \ldots, 0.01}_{5}, \underbrace{0, \ldots, 0}_{p-15})$ or called sparse data, i.e., too many non-significant predictors. The models of interest in this paper are:

(1) multiple linear regression model following the form:

$$\mathbf{y} = \mathbf{X}^{\mathbf{T}}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad (2.12)$$

where $\mathbf{y} = (y_1, y_2, ..., y_n)^T$ is a $n \times 1$ of dependent variable, $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_p)^T$ is a $p \times 1$ vector of unknow parameters, $\mathbf{X}$ for $i = 1, 2, \ldots, n$ is a $n \times p$ vector of independent variable ,and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)^T$ is a $n \times 1$ vector of random errors. The random errors $\boldsymbol{\varepsilon}$ were generated from normal and positively skew distributions: (i) standard normal distribution or $N(0,1)$, (ii) $\text{Gamma}(\alpha,1)$, $\alpha = 16, 4, 0.25, 0.016, 0.004$ corresponding to coefficient of skewness (*sk*) which is equal to $0.5, 1, 4, 5$, and 30, respectively, (iii) $\text{Lognormal}(0,\sigma)$, where $\sigma = 0.3143, 0.5514, 0.9202, 1.4733$ corresponding to coefficients of skewness equal to $,1,2,5,30$ respectively, and (iv) Weibull($1,\lambda$), where $\lambda =$ 2.88, 1.5, 0.5999, 0.39 corresponding to coefficient of skewness which equals $0.2, 1, 5$, and 10, respectively.

(2) logistic regression model follows the form:

$$P\left(y=1\middle|\mathbf{x_i}\right)=\pi_i=\frac{\exp\left(\mathbf{x_i^T\beta}\right)}{1+\exp\left(\mathbf{x_i^T\beta}\right)}. \quad (2.13)$$

(3) Poisson regression model with mean parameter μ_i which can be defined as

$$\mu_i=\exp\left(\mathbf{x_i^T\beta}\right). \quad (2.14)$$

2.3 Performance criteria

Two criteria are employed in this paper. The prediction accuracy is evaluated by the mean of prediction mean square errors (mPMSE) defined as:

$$\text{mPMSE}_j=\frac{\sum_{j=1}^{1000}\text{PMSE}_j}{1,000}$$
$$=\frac{\sum_{j=1}^{1000}\left[\sum_{i=1}^{n}(y_i-\hat{y}_i)^2/n\right]_j}{1,000}, \quad (2.15)$$

where the number of replicates equals 1,000 in the Monte Carlo simulation. The variable selection performance is accessed using the false positive rate (FPR) and false negative rate (FNR). The FNR is the proportion of variables excluded from a model, although their coefficients are not zero; the FPR is the proportion of variables included in a model, but the true coefficients are zero. The formulas are the following:

$$FPR=\frac{\left\{j=0,\ldots,p;\beta_j=0,\hat{\beta}_j\neq 0\right\}}{\left\{j=0,\ldots,p;\ \hat{\beta}_j\neq 0\right\}} \quad (2.16)$$

and

$$FNR=\frac{\left\{j=0,\ldots,p;\beta_j\neq 0,\hat{\beta}_j=0\right\}}{\left\{j=0,\ldots,p;\ \hat{\beta}_j=0\right\}}. \quad (2.17)$$

Even though there are three criteria, the mPMSE is used as the main one to decide which method is the best because regression analysis is often interested in making predictions.

3. Results and Discussion

It should be noted that FPR and FNR results were not presented in detail, but rather as an overall picture, and that the Ridge regression method does not involve variable selection.

First, the efficiency of ALASSO and AEnet methods was studied in multiple linear models with normal and positively skewed distributions: gamma, log-normal, Weibull error terms, Poisson, and logistic regression models. For overall prediction, both the ALASSO and AEnet give the lowest mPMSE when the adaptive weight was constructed by the Ridge estimator. All mPMSE values for a different configuration of n and p are represented graphically in Figs. 1. and 2.

While Ridge estimators produce the best result, the adaptive weight from the Enet estimator is the second-best, followed by the LASSO estimator. Only the example of a logistic regression model with constant correlation structures will be discussed here; other models with varying correlations will not be given because the general conclusions will be the same.

The effectiveness of penalized regression coefficient estimation using five distinct methods will be presented here. It is noted that the Ridge estimation was used to determine the adaptive weight on the ALASSO and AEnet. The results will be divided into three parts based on the models.

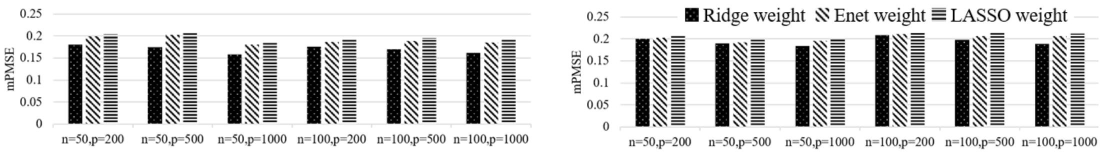

Fig. 1. mPMSE obtained from ALASSO with three different adaptive weights for constant $r = 0.5$ (left) and $r = 0.9$ (right).

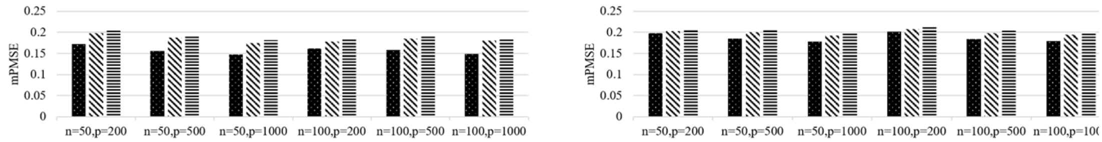

Fig. 2. mPMSE obtained from AEnet estimator with different adaptive weights for constant $r = 0.5$ (left) and $r = 0.9$ (right).

3.1 Linear regression model

The AEnet gives the lowest mPMSE among all others in the multiple linear regression model with normal error terms in all correlation structures. Generally, ALASSO outperforms LASSO and Ridge in terms of prediction. While ALASSO outperforms Enet in most cases, Enet is better at making predictions than LASSO. The result is summarized in Table 2.

Table 2. Summary of the results of the best prediction accuracy and variable selection performance in multiple linear, Poisson, and logistic regression models

Model	Distribution of errors	Level of skewness	Coefficient of skewness	Prediction criterion: mPMSE, Correlation structure (C T HT)	Variable selection: FPR, Correlation structure (C T HT)	Variable selection: FNR, Correlation structure (C T HT)
	Normal		0	AEnet	Enet/AEnet	ALASSO
Multiple linear regression	Gamma	Low	0.5	AEnet	AEnet	LASSO
			1			
		Moderate	4		Enet	
			5	LASSO/Enet/ALASSO/AEnet		ALASSO
		High	30			
	Log-normal	Low	1	LASSO/Enet/ALASSO/AEnet	Enet	ALASSO
		Moderate	2	AEnet		LASSO
			5		AEnet	
		High	30			
	Weibull	Low	0.2	LASSO/Enet/ALASSO/AEnet	Enet	ALASSO
		Moderate	1	AEnet		LASSO
			5		AEnet	
		High	10			
Poisson				AEnet	Enet/AEnet	LASSO
Logistic				AEnet	Enet/AEnet	LASSO

Note: C = constant, T = toeplitz, HT = hub toeplitz

For fixed n and $r = 0.9$ with a constant and Hub Toeplitz correlation structure, the performance of Ridge, ALASSO, and AEnet increases (mPMSEs decrease) as p increases. Furthermore, the ALASSO and the AEnet have higher prediction performance when n is much larger than p. All results are illustrated in Figs. 2 and 3.

For the multiple linear regression model with positive skew distribution error terms, the following are the main outcomes:

I. In a multiple linear regression model with a gamma error term, the AEnet gives the lowest mPMSE at low to moderate skewness. In terms of mPMSE, ALASSO often outperforms Enet, LASSO, and Ridge. Except for the Hub Toeplitz correlation, the performance of LASSO, Enet, and ALASSO increases as *p* increases for fixed *r* and *n* with a moderate to high level of skewness.

II. At low levels of skewness, LASSO, Enet, ALASSO, and AEnet all do about the same in the multiple linear regression model with log-normal error components. For levels of moderate to high skewness, AEnet frequently provided the lowest mPMSE. ALASSO often outperforms Enet, LASSO, and Ridge. Ridge fared better than LASSO and Enet at moderate to high skewness levels.

III. When a multiple linear regression model with a Weibull error term is considered, LASSO, Enet, ALASSO, and AEnet provide comparable prediction performance at the low level. At moderate to high levels of skewness, it is found that AEnet often gives the lowest mPMSE. In details, ALASSO outperforms Enet, LASSO, and Ridge, respectively. At moderate to high levels of skewness, Ridge outperforms LASSO and Enet. From Figs. 3–5, the performance of LASSO, ALASSO, Enet, and AEnet is improved when *n* increases.

3.2 Poisson regression model

Figs. 3–5 show that the AEnet appears to have the lowest mPMSE among all the five methods considered. It is discovered that the ALASSO outperforms the LASSO and Ridge in prediction. Except for the Hub Toeplitz correlation, the prediction performance of all estimators improves as *r* increases or the correlation becomes stronger for fixed *n* and *p*. Considering incorrect variable selection, it is found that AEnet and Enet give the lowest FPR. In terms of FNR, LASSO gives the lowest, whereas AEnet gives the highest. The results of mPMSE for fixed *r* and *n* are like those of the multiple linear regression model with a normal error term. The conclusions are summarized in Table. 2.

3.3 Logistic regression model

The results of prediction performance shown in Figs. 6-8 are similar to those from the Poisson regression model in terms of the lowest mPMSE. When *r* and *n* are kept the same, ALASSO and AEnet estimators do a better job of predicting as *p* increases. The ALASSO and AEnet tend to make better predictions when *n* is very large than when *n* is close to *p*. Overall, for fixed *r* and *n* or fixed *r* and *p*, the results of prediction performance are similar to those of the multiple linear with normal error term and Poisson regression models.

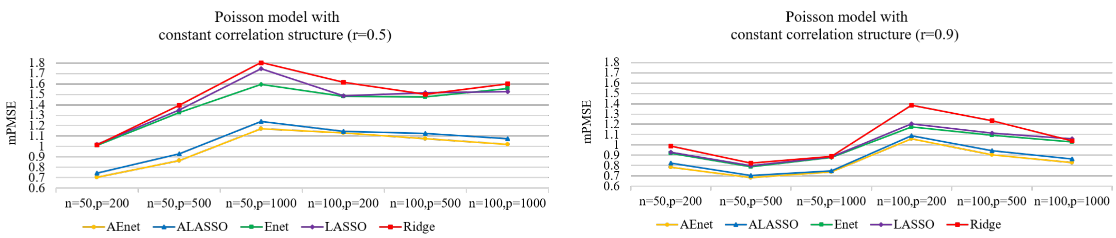

Fig. 3. Comparisons of mPMSEs for the constant correlation structure in Poisson regression models.

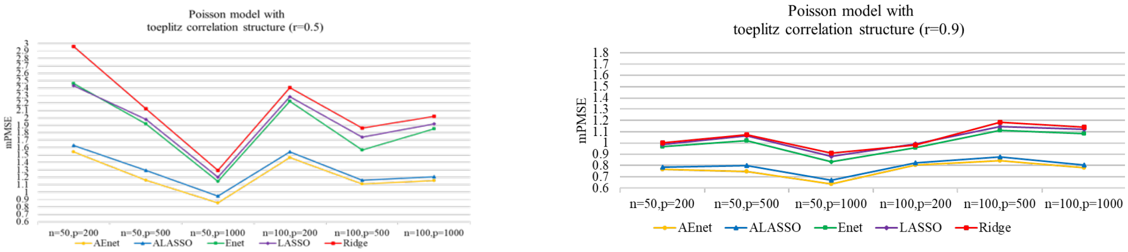

Fig. 4. Comparisons of mPMSEs for the Toeplitz correlation structure in Poisson regression models.

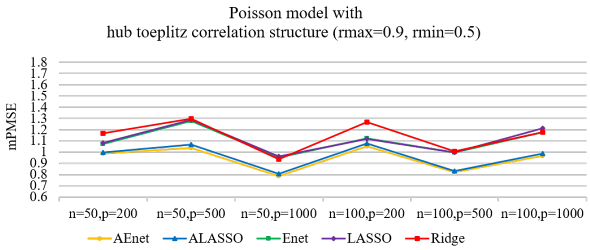

Fig. 5. Comparisons of mPMSE for the Hub Toeplitz correlation structure in Poisson regression models.

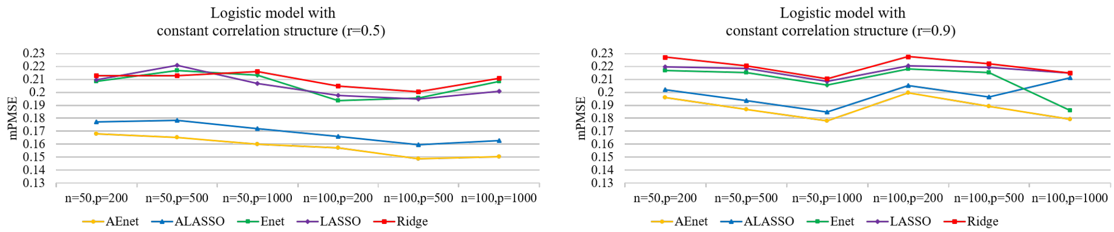

Fig. 6. Comparisons of mPMSEs for the constant correlation structure in logistic regression models.

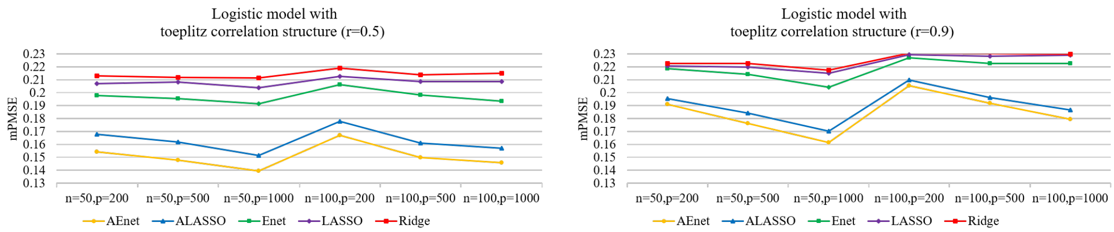

Fig. 7. Comparisons of mPMSEs for the Toeplitz correlation structure in logistic regression models.

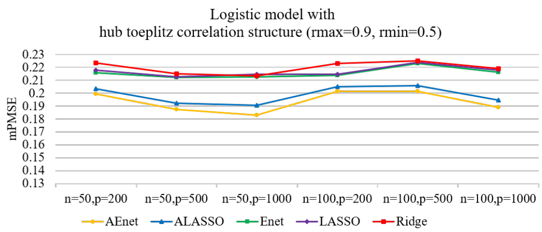

Fig. 8. Comparisons of mPMSEs for the Hub Toeplitz correlation structure in logistic regression models.

4. Real Data Application

For the real data sets, it is noted that the Ridge estimator was used to find adaptive weights for ALASSO and AEnet.

4.1 Multiple linear regression model

The data in this example are:

I. Individuals from a genetically homogeneous sample who produced riboflavin in Bacillus subtilis were the dependent variable, and measures of riboflavin logarithm production rate were the independent variables [18]. The Kolmogorov-Smirnov test was used to examine residuals for normality, and it was found that they were not. According to Table 4, the AEnet had better prediction performance than the others.

II. From 123 patients in the CHEMORES cohort who underwent complete surgical resection, a genomic collection of lung cancer was obtained [19]. The dependent variable was the time of disease-free survival, and the independent variables were 940 quantitative variables. After being checked for normality, the Kolmogorov-Smirnov test residuals were found to be so. Table 5 shows that the AEnet estimator outperformed the others in terms of prediction accuracy.

4.2 Poisson regression model

In this case, the data came from the dataset for the Software Engineering Teamwork Assessment in an Educational Setting [20]. The dependent variable, which was based on 74 observations of the teams' teamwork, was a count of the number of students who learned about software engineering teamwork. It had 81 independent variables, such as the number of women who attended, how long the meetings lasted, and how long they lasted on average. Table 6 demonstrates that, in terms of prediction, AEnet performs better than the others.

Table 4. Comparison of mPMSEs among the five methods using a multiple linear regression model with non-normal residuals.

Ridge	LASSO	Enet	ALASSO	AEnet
0.0623	0.0946	0.0616	0.0522	0.0510

Table 5. Comparison of mPMSEs among the five methods using a multiple linear regression model with normal residuals.

Ridge	LASSO	Enet	ALASSO	AEnet
3.7908	4.2756	3.8133	3.5719	3.0859

Table 6. Comparison of mPMSEs among the five methods using a Poisson regression.

Ridge	LASSO	Enet	ALASSO	AEnet
0.3478	0.1205	0.1165	0.1138	0.1105

Table 7. Comparison of mPMSEs among the five methods using a logistic regression.

Ridge	LASSO	Enet	ALASSO	AEnet
0.0154	0.0095	0.0085	0.0097	0.0083

4.3 Logistic regression model

The data in this example were gene expression measurements of 72 leukemia patients, 47 with acute lymphoblastic leukemia (ALL), and 25 with acute myeloid leukemia (AML) [21]. The dependent variable is a binary variable with only two possible values, and the independent variable is one of 3571 quantitative variables. In terms of making predictions, Table 7 shows that the AEnet did the best, followed by the Enet, LASSO, ALASSO, and Ridge estimators.

5. Conclusion

Using the Ridge estimator to construct the adaptive weights produces the most accurate predictions for ALASSO and AEnet. In terms of prediction accuracy, from simulation studies, AEnet tends to outperform the other four approaches in multiple linear with normal error terms, Poisson, and logistic regression models. In certain instances, particularly for regression models with non-normal errors, LASSO, ALASSO, Enet, and AEnet all give comparable prediction accuracy. However, correlation structures have a limited effect on the mPMSE.

Different strategies are required to obtain the lowest false positive rate (FPR) and false negative rate (FNR) given the objective. Enet or AEnet is required to obtain the lowest FPR, while LASSO or ALASSO will produce the lowest FNR.

Acknowledgements

The authors express their gratitude to the editor and reviewers for their valuable comments and suggestions that have helped improve this paper. Dr. Srisuradetchai's research is presently supported by the Thammasat University Research Unit in Theoretical and Computational Statistics. We also sincerely appreciate the financial assistance provided by Thammasat University.

References

[1] Raymond H, Douglas C, Geoffrey G, Timothy J. Generalized Linear Models with Application in Engineering and the Sciences. United States of America: A john wiley & sons; 2010.

[2] Algamal ZY, Lee MH. Adjusted adaptive lasso in high-dimensional Poisson regression. Modern Applied Science 2015;9(4):170-7.

[3] Oyeyemi GM, Ogunjobi EO, Folorunsho Al. On performance of shrinkage methods - a Monte Carlo study. International Journal of Statistics and Applications 2015;5(2):72-6.

[4] Tibshirani R. Regression shrinkage and selection via the lasso. Journal of Royal Statistical Society. Series B (Methodological) 1996;58(1):267-88.

[5] Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. Technometrics 1970;12(1):55-67.

[6] Zou H, Hastie T. Regularization and variable selection via the elastic net. J Royal Statistical Society B 2005;67(2):301-20.

[7] Zou H. The adaptive lasso and its oracle properties. Journal of the American Statistic Association 2006;101(476): 1418-29.

[8] Jiratchayut K, Bumrungsup C. A study of adaptive elastic net estimators with different adaptive weights. Thammasat International Journal Science and Technology 2015;20(3):1-7.

[9] Zou H, Zhang T. On the adaptive elastic-net with diverging number of parameters. The Annals of Statistic 2009;37(4):1733-51.

[10] Choosawat O, Reangsephet O, Srisuradetchai P, Lisawadi S. Performance comparison of penalized regression methods in Poisson regression under high-dimensional sparse data with multicollinearity. Thailand Statistician 2020;18(3):306-18.

[11] Sarakor T, Kulvanich N. Comparing the prediction accuracy and subset selection performances of stepwise, lasso, elastic net and adaptive lasso for small and sparse signals. Proceedings of the Rmutto Research Conference 2008:327-31.

[12] Lisawadi S, Watcharasatian W, Tulyanitikul B. Performance comparison of penalized regression method in logistic regression for high-dimensional sparse data with Multicollinearity. Thai Journal of Statistics and Technology 2020;9(6):761-72.

[13] Neammai J, Lisawadi S. Performance comparison of penalized linear regression method with non-normal random error distributed for high-dimensional sparse data and multicollinearity [Master's thesis]. Pathum Thani: Thammasat University; 2020.

[14] Rungsaranon B, Araveeporn A. comparing methods of parameter estimation with penalized regression analysis under high-dimensional data. Thai Science and Technology Journal 2020;28(8):1347-58.

[15] Porndumnernsawat P, Hirunkasi kannigarh. Performance comparison of lasso methods for parameter estimation in high dimensional linear regression [Master's thesis]. Bangkok: Silpakorn University; 2017.

[16] Hossain S, Ahmed S. Shrinkage estimation and selection for logistic regression model, CRM Proc. Contemporary Mathematics 2014;622:159-76.

[17] Algamal ZY, Lee MH, AI-Fakih AM, Aziz M. High-dimensional QSAR prediction of anticancer potency of imidazo[4,5-b] pyridine derivatives using adjusted adaptive lasso. J Chemometrics 2015;29:547-56.

[18] PLoS ONE [Internet]. [cited 2020 Sep 17]. Available from: https://doi.org/ 10.1371/journal.pone.0183518

[19] Genomic data [Internet]. [cited 9 Nov 2020]. Available from https://github.com/ jedazard/PRIMsrc

[20] Data for Software Engineering Teamwork Assessment in Education Setting Data Set [Internet]. [cite 11 May 2021]. Available from https://archive.ics. uci.edu/ml/datasets/Data+for+Software+ Engineering+Teamwork+Assessment+in +Education+Setting

[21] Leukemia data [Internet]. [cite 24 Sep 2020]. Available from https://web.stanford.edu/~hastie/CASI_files/DATA/leukemia.html