Available online at http://iorajournal.org/indx.php/orics/index

Operations Research: International Conference Series

e-ISSN: 2722-0974
p-ISSN: 2723-1739

Vol. 3, No. 1, pp. 8-13, 2022

Comparing the Performance of Prediction Model of Ridge and Elastic Net in Correlated Dataset

Richy Marcelino Bastiaan[1*], Deiby Tineke Salaki[2] and Djoni Hatidja[3]

[1,2,3]Department of Mathematics, Sam Ratulangi University, Manado, Indonesia

**Corresponding author email: richybastiaan103@student.unsrat.ac.id*

Abstract

Multicollinearity refers to a condition where high correlation between independent variables in linear regression model occurs. In this case, using ordinary least squares (OLS) leads to unstable model. Some penalized regression approaches such as ridge and elastic-net regression can be applied to overcome the problem. Penalized regression estimates model by adding a constrain on the size of parameter regression. In this study, simulation dataset is generated, comprised of 100 observation and 95 independent variables with high correlation. This empirical study shows that elastic-net method outperforms the ridge regression and OLS. In correlated dataset, the OLS is failed to produce a prediction model based on mean squared error (MSE).

Keywords: Elastic-net, Multicollinearity, Ordinary Least Square, Penalized Regression, Ridge Regression

1. Introduction

Regression is a statistical analysis used for describing models that estimate the relationships among variables. Linear regression models study the relationship between a single dependent variable Y and one or more independent variables, denoted by X (Bangdiwala 2018; Gurunathan & Kim 2016). In regression analysis, if regression using single independent variables is called simple linear regression while the analysis using more than one independent variable is called Multiple Linear Regression. In regression, to find the estimation parameter using ordinary least square method.

Regression also had some problems in the analysis, the one of them is if the multicollinearity is existed. Multicollinearity appears when two or more independent variables in the regression model are correlated. One of the methods that can be used to detect multicollinearity is using variance inflation factor (VIF). If the VIF value is more than 10, then the data have some multicollinearity problem (Wilcox & Keselman 2012). To solve this multicollinearity problem, we can use the penalized regression method.

Penalized regression is a method that used a penalized term in searching for estimation parameters so that the shrinkage can be controlled. Penalized regression has several methods, including ridge regression, LASSO (Least Absolute Shrinkage and Selection Operator) and elastic-net. This research was conducted to prove that the penalized regression method: ridge regression and elastic-net can outperform the problem of multicollinearity based on the comparison of their MSE value with the MSE value of the ordinary least square.

2. Literature Review

2.1.Multiple Linear Regression

Multiple linear regression is regression analysis method that predicted the value of the dependent variable based on the value of the independent variable with the number of independent variables more than one variable (Schneider, et al., 2010). Here is the formula for multiple regression:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ij} + \varepsilon_i \tag{1}$$

Y : Dependent Variable
X : Independent Variables
i : Number of Observations

j : Number of Variables
β : Regression Coefficients/Estimation Parameters ($j = 0,1,2,\dots,k$)
ε_i : Error Term

Ordinary least square (OLS) is generally used to estimate the parameter ($\hat{\beta}$) of equation (1) (Long & Ervin, 2000; Raftery, et al., 1997) by following the formula as follows:

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (2)$$

In this approach there are many assumptions that must be fulfilled. If one or more assumption is violated, then the model in hand is no more reliable and also is not acceptable in estimating the population parameters. According to Mebane (1992) the assumptions that must be fulfilled in multiple linear regression consist of follows:

1. The regression model is linear in parameter
2. The average value of the error is zero
3. The variance of the *error* is constant (Homoscedastic)
4. There is no autocorrelation on the *error*
5. Multicollinearity does not occur in the independent variables
6. *Errors* are normally distributed

2.2. Multicollinearity

Multicollinearity is a problem caused by a fairly high correlation between several independent variables from the data (Filler, 2009). Correlation is an association between two regular random variables with the same or opposite directions. According to Narayan (2005). multicollinearity can have some impacts including:

1. the variance and covariance of the least square's estimator become larger,
2. A large standard error causes the confidence interval for the parameter to be larger,
3. high R^2value but some t-statistical values are not real.

Multicollinearity can be detected by finding the variance inflation factors (VIF) value and tolerance (TOL) value. If the VIF value is about 5-10 or the tolerance value is less than 0.1, it can be said that multicollinearity is exist. Formulas for VIF and tolerance follow equation (3) and (4) (Widmann, et al., 2019).

$$VIF_j = \frac{1}{1-R_j^2} \quad (3)$$

$$TOL_j = \frac{1}{VIF_j} \quad (4)$$

2.3. Penalized Regression

Penalized regression is used when there are an excessive number of independent variables in the regression or high-dimensional problems (Alfons, et al., 2009; Xie& Huang, 2009). Penalized regression keeps all the predictor variable in the model but constrain (regularize) the regression coefficients by shrinking them toward zero. Penalized regression using a penalty value to control the shrinkage of the estimated parameter value that caused due to the OLS method. Penalized regression has several methods including ridge regression, LASSO, and elastic-net. The following is the formula used to find penalized regression:

$$\hat{\beta} = \arg min_{\beta}\left[\sum_{i=1}^{n}\left(Y_i - \beta_0 - \sum_{j=1}^{p}\beta_j X_{ji}\right) + \lambda\sum_{j=1}^{p}|\beta_j|^q\right] \quad (5)$$

For the formula $\sum_{i=1}^{n}\left(Y_i - \beta_0 - \sum_{j=1}^{p}\beta_j X_{ji}\right)$ is an equation for ordinary least square (OLS) and $\lambda\sum_{j=1}^{p}|\beta_j|^q$is a penalty value that given by penalized regression. The lambda value (λ) is a parameter for determine the shrinkage that occurs for the equation. For ridge using Euclidean distance with q = 2 in the equation (5) and for lasso using Manhattan distance with q = 1.

2.4. Ridge Regression

Ridge regression is perfect if there are various predictors, all with non-zero coefficients and collect from a normal distribution (Hoerl& Kennard, 1970). In specific, it carries out well with many predictors each having small outcome and prevents coefficients of linear regression models with many correlated predictors from being poorly determined and exhibiting high variance. Ridge regression shrinks the coefficients of correlated predictors equally towards zero. The following is the estimation formula for Ridge Regression:

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum\nolimits_{j=1}^{p}\beta_j X_{ij}\right)^2 + \lambda\sum\nolimits_{j=1}^{p}{\beta_j}^2 \quad (6)$$

The lambda (λ) is a parameter that will determine the shrinkage that will happen in the formula. if the value of $\lambda = 0$ then the penalty of ridge regression will not affect the ordinary least square process, but if the value of $\lambda = 1 \dots \infty$ then the penalty from the ridge regression will apply in the process of the ordinary least squares.

2.5. Elastic Net

Elastic-net is a penalized regression method that combining the penalty value between ridge regression and LASSO. According to Zou & Hastie (2005) this elastic net was formed to cover the weakness of ridge and lasso because the two formulas have gaps including:

1. When number of variables is more than number of observations, then LASSO only selects the number of observation variables included in the model.
2. If there is a set of variables with a high correlation, then LASSO just randomly chooses one of the variables.
3. When number of variables is less than number of observations, LASSO performance will be dominated by ridge regression.

Here is the penalty formula from elastic-net:

$$\sum_{j=1}^{p}\left[a\left|\beta_j\right| + (1-a){\beta_j}^2\right] \tag{7}$$

Elastic-net regularization can be used to analyze dataset in correlated condition as done by ridge regression and simultaneously doing selection variables such as LASSO (Vapnik, 1999; Romberg &Saffran, 2010). The methods minimize the error of model with certain penalty as follows:

$$\sum_{j=1}^{p}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j X_{ij}\right)^2 + \lambda_2\sum_{j=1}^{p}{\beta_j}^2 + \lambda_1\sum_{j=1}^{p}\left|\beta_j\right| \tag{8}$$

Here λ_2 is a shrinkage parameter for ridge regression and λ_1 is a shrinkage parameter of LASSO. For the Equation can be written as this following formula:

$$\sum_{j=1}^{p}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j X_{ij}\right)^2 + \lambda\left[(1-\alpha)\sum_{j=1}^{p}{\beta_j}^2 + \alpha\sum_{j=1}^{p}\left|\beta_j\right|\right] \tag{9}$$

Where $\alpha = \frac{\lambda_1}{\lambda_1+\lambda_2}, 0 \leq \alpha \leq 1$. α is combination of shrinkage parameter between ridge regression and lasso. If the a-value is 0 then the penalty that will be used is penalty of ridge regression, but if the a-value is 1 then the penalty that will be used is penalty of LASSO.

3. Research Methodology

The research employs simulation data which is generated by using library MASS in RStudio software. The dataset includes 100 observations and 95 independent variables which are further set as matrix X. The matrix X is adjusted in high correlated condition with correlation value equals to 0.95. Response variable ($\boldsymbol{y}$), is generated by using formula $y = X\beta + \varepsilon$. Here some items of $\boldsymbol{\beta}$ set as 1 and others as 0. The error term ε is generated by following the normal distribution with $\mu = 0$ and$\sigma^2 = 0.5$. Next, the cross-validation procedure is applied to build model using the training and validate using testing dataset for each of ridge regression and elastic-net analysis. The training dataset are employed to calculate the lambda value for penalty shrinkage ridge regression and elastic-net by using the library GLMNET. The prediction performance of the two methods is indicated by MSE of ridge regression and elastic-net.

4. Result And Discussion

In this research, a simulated data set with 100 observations and 95 independent variables will be analyzed. Multiple linear regression equation will be created based on the data in order to calculate the mean square error (MSE) value of the multiple linear regression. The regression equation can be written:

$$\hat{y} = -0.04095 - 1.06926X_1 + 3.21968X_2 + 1.66821X_3 - 1.80803X_4 - 0.60241X_5 + \cdots \\ -0.90973X_{95} \tag{10}$$

The MSE value resulted equals to 7.67×10^{64} which is calculated based on the multiple linear regression equation (10) using OLS by using RStudio. The VIF results and tolerance will next be analyzed to see if the model has multicollinearity. The results of VIF and tolerance of all data using R studio are as shown in Figure 1.

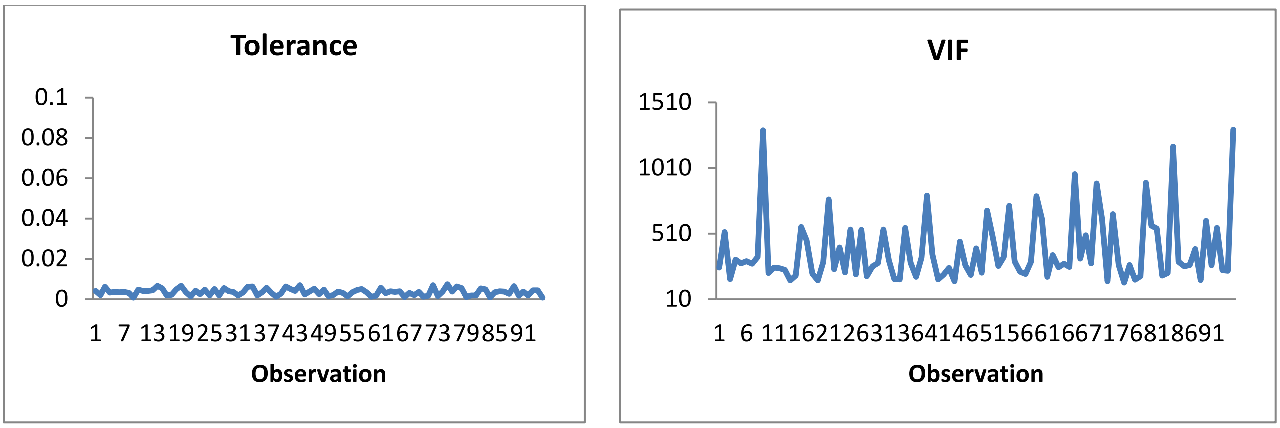


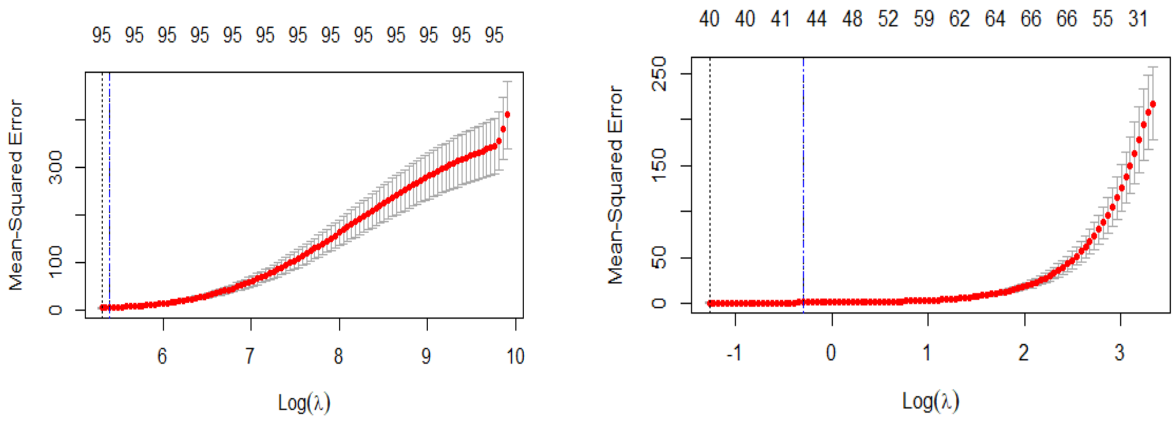
Figure 1. The Value of Tolerance and VIF

It can be seen from the Figure 1, VIF and tolerance value of the previous multiple linear regression model that several variables have multicollinearity issues. The data then be analyzed using ridge regression and elastic net to solve the problem, with determine the MSE and prediction value of ridge regression and elastic-net.

4.1. Penalized Regression

4.2.1. Ridge Regression

Using the cross-validation method model is build by ridge regression. The lambda of the ridge regression is calculated using training dataset and the predicted and MSE of ridge regression is calculated using testing datasets. The value of alpha (α) in the ridge regression formula is set to. Lambda.1se is used as the best lambda to determine best predictive model. Plot of MSE is displayed in Figure 2 (a). At the best lambda, that is $\log(161.1) = 5.08$, as indicated by blue dotted vertical line in each of the figure, the minimum MSE is reached.

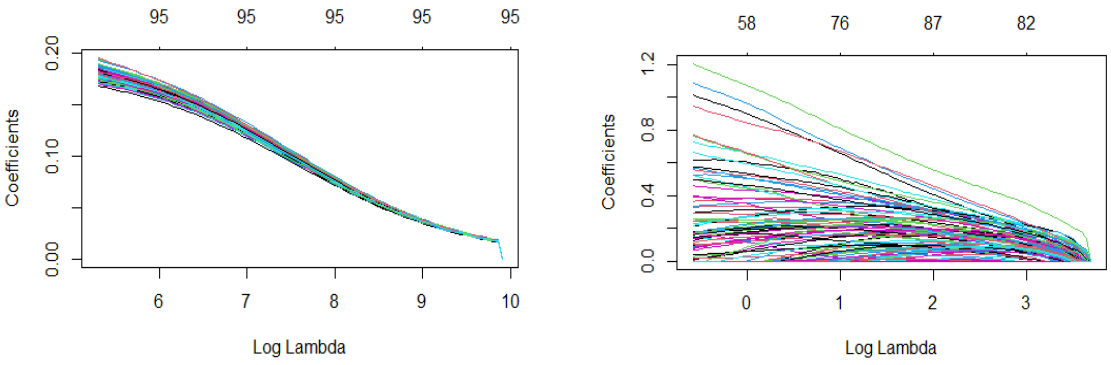


(a) (b)

Figure 2. Plot of MSE for Ridge (a) and Elastic Net (b)

4.2.2. Elastic net

Elastic-net is employed by using cross-validation with the same training and testing data set as did in ridge regression. In this method the value of alpha (α) is set to 0.5. Elastic net basically combined the penalty of ridge regression and LASSO. The best lambda is reached at 0.74, where $\log(0.74) = -0.3$, which is indicated by dotted vertical blue in Figure 2(b). The minimum MSE of the elastic net is produced at the point.



(a) (b)

Figure 3. Plot of coefficients for ridge regression (a) and Elastic Net (b)

Figure 3 displays plot of coefficient produced by ridge regression (a) and elastic net (b) in dependence of log lambda. The figures show that coefficient values decrease along with log lambda as the effect of penalization based on tunning parameter. In Figure 3(a), the value of coefficients reduced to almost zero as the lambda increase to infinity. On the other hand, in the other approach, coefficient can reduce to zero.

4.2. Comparison the Prediction and MSE value of Ridge regression and Elastic-net

Based on the results that have been found before, the lambda reflects the size of the shrinkage penalty that will be applied to the testing dataset from cross validation to determine the predicted value of ridge regression and elastic-net. Figure 4 displays simultaneously the plot of actual response and both predicted values resulted by ridge regression and elastic-net.

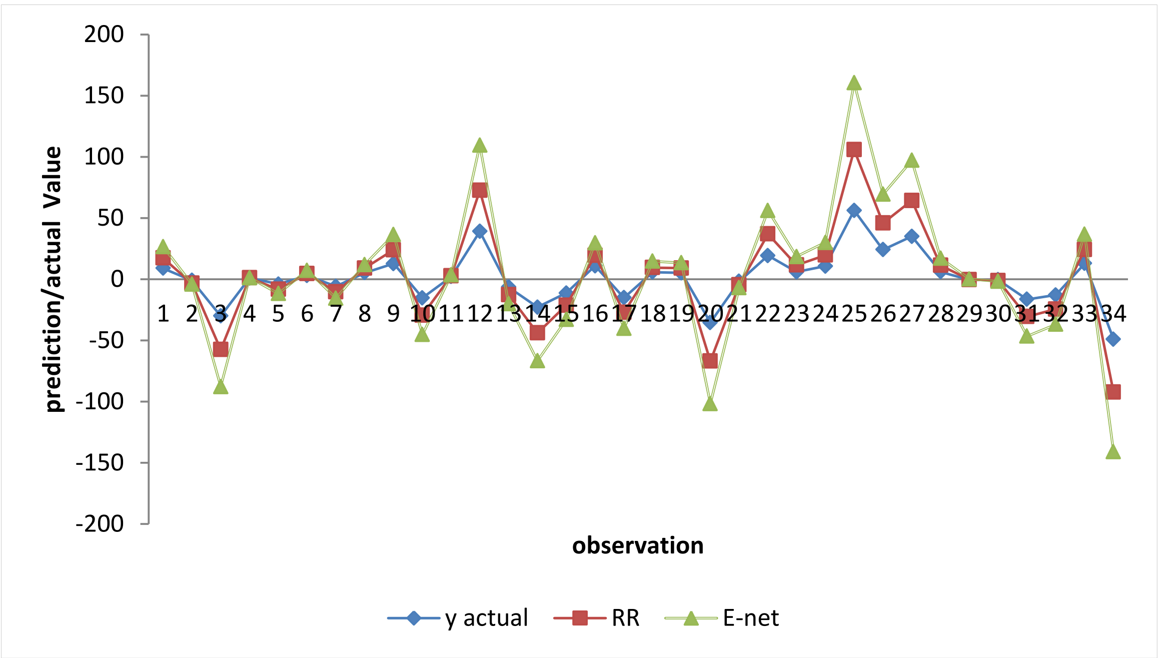


Figure 4. Plot of Predictor Value Ridge Regression and Elastic-net

According to figure, it can be seen that the pattern of predicted value of ridge regression is more coincide with actual than and elastic-net. On the side, Table 1 shows that MSE of ridge regression and elastic value are much lower than OLS. It can be concluded that prediction model of Ridge and Elastic net are better than one of OLS. The table shows that MSE of elastic-net is 0.83 which is superior to ridge regression with the value of 6.50. It can be said that model prediction resulted from penalized regression is better than from OLS. In high correlated data, the OLS is failed to produce a prediction model as the value of MSE is too large.

Table 1. MSE Value

Method	MSE
Ordinary Least Square	7.67×10^{64}
Ridge Regression	6.50
Elastic-Net	0.83

5. Conclusion

Based on the research conducted, the following two conclusions were obtained: Based on the distribution of predicted values, it is proven that the ridge regression line graph is more regular and closer to the Actual Y line compared to elastic-net. Based on the MSE value that has been obtained, it is proven that elastic-net method out performs the ridge regression and OLS. In high correlated data, the OLS is failed to produce a prediction model.

References

Alfons, A., Croux, C., & Gelper, S. (2013). Sparse least trimmed squares regression for analyzing high-dimensional large data sets. *The Annals of Applied Statistics*, 226-248.

Bangdiwala, S. I. (2018). Regression: simple linear. *International journal of injury control and safety promotion*, *25*(1), 113-115.

Filler, M. G. (2009). A Second Course in Regression Analysis as Applied to Valuation and Lost Profits. *Business Valuation Review*, *28*(2), 67-87.

Gurunathan, S., & Kim, J. H. (2016). Synthesis, toxicity, biocompatibility, and biomedical applications of graphene and graphene-related materials. *International journal of nanomedicine*, *11*, 1927.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55-67.

Long, J. S., & Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician*, *54*(3), 217-224.

Mebane, W. R. (1992). Analyzing the effects of local government fiscal activity I: Sampling model and basic econometrics. *Political Analysis*, *4*, 1-40.

Narayan, P. K. (2005). The saving and investment nexus for China: evidence from cointegration tests. *Applied economics*, *37*(17), 1979-1990.

Raftery, A. E., Madigan, D., & Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, *92*(437), 179-191.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(6), 906-914.

Schneider, A., Hommel, G., & Blettner, M. (2010). Linear regression analysis: part 14 of a series on evaluation of scientific publications. *Deutsches Ärzteblatt International*, *107*(44), 776.

Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE transactions on neural networks*, *10*(5), 988-999.

Widmann, D., Lindsten, F., & Zachariah, D. (2019). Calibration tests in multi-class classification: A unifying framework. *Advances in Neural Information Processing Systems*, *32*.

Wilcox, R. R., & Keselman, H. J. (2012). Modern regression methods that can substantially increase power and provide a more accurate understanding of associations. *European journal of personality*, *26*(3), 165-174.

Xie, H., & Huang, J. (2009). SCAD-penalized regression in high-dimensional partially linear models. *The Annals of Statistics*, *37*(2), 673-696.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, *67*(2), 301-320.