

# Adjusted Adaptive LASSO in High-Dimensional Poisson Regression Model

Zakariya Y. Algama<sup>1</sup> & Muhammad H. Lee<sup>1</sup>

<sup>1</sup> Department of Mathematical Sciences, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia

Correspondence: Muhammad H. Lee, Department of Mathematical Sciences, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia. Tel: 60-7553-236. Fax: 60-7556- 6162. E-mail: mhl@utm.my

Received: November 4, 2014 Accepted: November 10, 2014 Online Published: January 11, 2015

doi:10.5539/mas.v9n4p170

URL: <http://dx.doi.org/10.5539/mas.v9n4p170>

## Abstract

The LASSO has been widely studied and used in many applications, but it not shown oracle properties. Depending on a consistent initial parameters vector, an adaptive LASSO showed oracle properties, which it is consistent in variable selection and asymptotically normal in coefficient estimation. In Poisson regression model, the usual adaptive LASSO using maximum likelihood coefficient estimators can result in very poor performance when there is multicollinearity. In this study, we proposed an adjusting of the adaptive LASSO to take into account the maximum likelihood standard errors of the coefficient parameters. The performance of the adaptive LASSO was demonstrated through simulation and real data. Our simulation and real data results show that adaptive LASSO has advantage in terms of both prediction and variable selection comparing with other existing adaptive penalized methods when the explanatory variables are highly correlated. Hence we can conclude that adaptive LASSO is a reliable adaptive penalized method in a Poisson regression model.

**Keywords:** high dimensional, penalization, Poisson regression, LASSO, Adaptive LASSO

## 1. Introduction

With the advancement of technologies, massive amount of data with increasing dimensions have been generated in many areas such as genetics, medical, economic and social sciences. The expansion of the data is in two dimensions: the number of variables and the number of observations. Such high-dimensional data has posed new challenges to statistical analysis, because a lot of classically statistical methods do not automatically apply into these datasets, for example, the curse of dimensionality makes many classical regression models, such as Poisson regression model, ineffective, because statistical issues associated with modeling high-dimensional data include model overfitting, estimation instability, computational difficulty (Pourahmadi, 2013).

How to reduce the dimensionality has been an important research question in statistical application. One way to handle the high dimensional data is to perform data reduction. To do this, various penalized methods have been proposed begin by ridge penalty (Hoerl & Kennard, 1970). It estimates the regression coefficients through  $\ell_2$ -norm penalty. It is well-known that ridge regression shrinks the coefficients of correlated predictor variables toward each other, allowing them to borrow strength from each other (Friedman, Hastie, & Tibshirani, 2010). LASSO (least absolute shrinkage and selection operator), introduced by Tibshirani (1996), is another most frequently used penalized method. LASSO imposes the  $\ell_1$ -norm penalty to the residual sum of square. Because of the  $\ell_1$ -norm property, LASSO can perform variable selection by assigning some explanatory variable coefficients to zero. For this reason, LASSO gets its popularity in high dimensional data. Despite the advantage of the LASSO, it has some shortcomings. First, it cannot select more explanatory variables than the number of observations. Second, when there is multicollinearity, LASSO tray to select one variable among correlated variables. Last, LASSO does not have the oracle properties, which were referred to that the probability of selecting the right set of explanatory variables (with nonzero coefficients) converged to one and that the estimators of the nonzero coefficients were asymptotically normal with the same means and covariances as if the zero coefficients were known in advance. To overcome the first two limitations, Zou & Hastie (2005) proposed the elastic net penalty, for which the penalty is a linear combination of  $\ell_1$ -norm and  $\ell_2$ -norm.

Related to the last limitation of LASSO, oracle properties, Fan & Li (2001) showed that LASSO did not enjoy the oracle properties because its model selection result could be inconsistent which it cannot guarantee a consistent identification of the true model. In addition, its estimators cannot be as efficient as the oracle. To

overcome this problem, Zou (2006) proposed the adaptive LASSO (ALASSO) in which adaptive weights are used for penalizing different coefficients in the  $\ell_1$ -norm penalty. He proved that, if the weights are chosen to be small amount for the important explanatory variables and large amount for the unimportant explanatory variables, then adaptive LASSO consistently can select the model. Zou (2006) presented that the performance of the ALASSO cannot outperform LASSO in term of prediction when the ordinary least square (OLS) estimators are variable. This is because it suffers from the highly correlated variables. Therefore, Wang et al. (2011) introduced the random LASSO which is depend on the bootstrap idea as a novel extension of LASSO to handle multicollinearity and to use the bootstrap estimates as initial weights in the second stage of their proposed method. Lu & Fine (2012) studied the adaptive LASSO under many model misspecifications in GLM, where OLS were used as initial weight. Using the LASSO estimates as an initial weight was proposed by Lian (2012). A study of controlling the false positive rate in adaptive LASSO was conducted by Sampson et al. (2013) where the OLS were used as initial weight. Recently, an adjusted adaptive LASSO using standard error was proposed by Qian & Yang (2013). They proposed to use the ratio of standard error of the OLS to the OLS coefficients as an initial weight to deal with multicollinearity. Furthermore, Zeng et al. (2013) proposed to use adaptive LASSO in count regression model, in particular in zero-inflated count data. They proposed to use the maximum likelihood estimators as an initial weight in low dimensional data.

In this study, the ratio of the standard error of the maximum likelihood (ML) estimator to the ML estimator was proposed as an initial weight in adaptive LASSO (RALASSO). Simulation and real data results show that the proposed method can perform better than both LASSO and adaptive LASSO using maximum likelihood estimator. The remainder of this paper organizes as follows. Section 2 covers the penalized Poisson regression methods. Description of the ALASSO and RALASSO are explained in sections 3 and 4, respectively. Sections 5 and 6 are devoted to simulation study and its results. The real data analysis results are covered by section 7. We end this paper with a conclusion in section 8.

## 2. Penalized Poisson Regression Model

Poisson regression models have received much attention in econometrics and medicine literature as model for describing count data that assume integer values corresponding to the number of events occurring in a given interval. The Poisson regression model is the most basic model, where the mean of the distribution is a function of the explanatory variables. This model has the defining characteristic that the conditional mean of the outcome is equal to the conditional variance (Algamal, 2012; Zeng et al., 2013). A procedure called penalization, which is always used in variables selection in high dimensional data, attaches a penalty term  $P_\lambda(\beta)$  to the log-likelihood function to get a better estimate of the prediction error by avoid overfitting. Recently, there is growing interest in applying the penalization method in the Poisson regression models. Friedman, Hastie, and Tibshirani (2010) developed an efficient algorithm for the estimation of a generalized linear model including Poisson regression with a convex penalty. Fan & Lv (2011) developed the methodologies of nonconcave penalization for generalized linear model. Hossain & Ahmed (2012) proposed Stein-type shrinkage estimator for the parameters of Poisson regression model. Wang et al. (2014) proposed a combination of minimax concave and ridge penalties and a combination of smoothly clipped absolute deviation and ridge penalties.

In Poisson regression model, the number of events  $y_i$  has a Poisson distribution with a conditional mean that depends on individual characteristics according to the structural model.

$$f(y_i | \mathbf{x}_i) = \frac{e^{-\theta_i} \theta_i^{y_i}}{y_i!}, \quad y_i = 0, 1, \dots; i = 1, 2, \dots, n \quad (1)$$

and the conditional mean parameter

$$\theta_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta}). \quad (2)$$

Under the assumption of independent observations, the log-likelihood function is given by

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^n \{y_i \mathbf{x}_i^T \boldsymbol{\beta} - \exp(\mathbf{x}_i^T \boldsymbol{\beta}) - \ln y_i!\}. \quad (3)$$

The penalized Poisson regression (PPR) is defined as

$$PPR = \ell(\boldsymbol{\beta}) + \lambda P(\boldsymbol{\beta}) \quad (4)$$

where  $\lambda$  is defined as a tuning parameter ( $\lambda \geq 0$ ). It controls the strength of shrinkage the explanatory variables, when  $\lambda$  takes larger value, more weight will be given to the penalty term. Since the value of  $\lambda$  is

depends on the data, it can be computed using cross-validation method (Y. Fan & Tang, 2013) (James, Witten, Hastie, & Tibshirani, 2013). Before solving the PPR, it is worth to make centering to the  $\mathbf{y}$  and standardization to  $\mathbf{x}_j$ , so that  $\sum_{i=1}^n y_i = 0$ ,  $\frac{1}{n} \sum_{i=1}^n x_{ij} = 0$ , and  $\sum_{i=1}^n x_{ij}^2 = 1$  for  $j = 1, 2, \dots, k$ , this is to make the intercept ( $\beta_0$ ) equals zero.

The LASSO for the Poisson regression model was originally proposed by Park & Hastie (2007). This technique is in some sense similar to ridge regression but it can shrink some coefficients to zero, and thus can implement variable selection. The LASSO method estimates the coefficients by minimizing the negative log-likelihood with the constraint that the sum of the absolute values of the model coefficients is bounded above by some positive number. The LASSO estimator is

$$\hat{\boldsymbol{\beta}}_{LASSO} = \arg \min_{\boldsymbol{\beta}} \left( -\ell(\boldsymbol{\beta}) + \lambda \sum_{j=1}^k |\boldsymbol{\beta}_j| \right) \quad (5)$$

where  $\lambda \geq 0$  is the tuning parameter. For large values of  $\lambda$ , Eq. (5) produces shrunken estimates of the  $\boldsymbol{\beta}$  and sets some variables to equal zero. LASSO can be efficiently solved by several methods, such as least angle regression algorithm (LARS) (Efron, Hastie, Johnstone, & Tibshirani, 2004) and the coordinate descent algorithm (Friedman et al., 2010).

Compared to the classical variable selection methods, LASSO has two advantages. Firstly, the selection process in LASSO is continuous which make the selection more stable than the subset selection. Secondly, LASSO is computational feasible in high dimensional generalized linear model (GLM). On the other hand, LASSO has three main drawbacks. First of all, LASSO selects at most  $n$  explanatory variables because of the nature of the convex optimization problem. In addition, LASSO cannot handle multicollinearity. When the pairwise correlations among a group of explanatory variables are very high, then LASSO tends to select only one explanatory variable from the whole group and does not take into account which one is selected. Lastly, LASSO lacks the oracle properties which is stated in Fan & Li (2001). Elastic net is a penalized method for variable selection, which is introduced by Zou & Hastie (2005) to deal with the first two drawbacks of LASSO. Elastic net tries to merge the  $\ell_2$ -norm and the  $\ell_1$ -norm penalties, by using ridge regression penalty to deal with high correlation problem while taking advantage of LASSO penalty in variable selection property.

### 2.1 Adaptive LASSO

According to Fan & Li (2001), a good penalty term should result in an estimator with three properties: unbiasedness, sparsity and continuity. Unbiasedness means the resulting estimator has no over penalization for large parameter to avoid unnecessary modeling biases. Furthermore, sparsity is another property that an estimator enjoys. In other words, the resulting estimator automatically set insignificant parameters to zero. Lastly, continuity is the third property, meaning that the resulting estimator is continuous in data in order to avoid instability in model prediction. Using the language of Fan & Li (2001), we called a penalty term enjoy oracle properties it has the following properties:

- It can identify the right subset model, i.e.,  $\{j : \hat{\boldsymbol{\beta}}_j \neq 0\} = \{j : \boldsymbol{\beta}_j \neq 0\}$ ,
- It has asymptotic normal distribution, i.e.,  $\sqrt{n}(\hat{\boldsymbol{\beta}}_A - \boldsymbol{\beta}_A) \rightarrow_d N(\mathbf{0}, \boldsymbol{\Sigma}^*)$ , where  $A = \{j : \hat{\boldsymbol{\beta}}_j \neq 0\}$  and  $\boldsymbol{\Sigma}^*$  represents the covariance matrix when the true subset model is known.

One of the main reasons for the LASSO not to be consistent, i.e., lacks the oracle property (J. Fan & Li, 2001) is that it equally penalized all the regression coefficients, which over-penalized the irrelevant explanatory variables leading it to be biased estimator. To alleviate this drawback, Zou (2006) proposed the adaptive LASSO in which adaptive weights are used for penalizing different coefficients in the  $\ell_1$ -norm penalty. The basic idea behind ALASSO is that, by assigning a higher weight to the small coefficients and lower weight to the large coefficients, it is possible to reduce the selection bias, therefore consistently can select the model. Furthermore, the ALASSO solution is continuous from its definition, which makes it enjoying oracle properties. The penalized likelihood using adaptive LASSO is defined by

$$\hat{\boldsymbol{\beta}}_{ALASSO} = \arg \min_{\boldsymbol{\beta}} \left\{ -\ell(\boldsymbol{\beta}) + \lambda \sum_{j=1}^k \mathbf{w}_j |\boldsymbol{\beta}_j| \right\}, \quad (6)$$

where  $\mathbf{w} = (w_1, \dots, w_p)^T$  is  $p \times 1$  data-driven weight vector. It depends on a

$\sqrt{n}$ -consistent initial values of  $\hat{\boldsymbol{\beta}}$  and  $w_j = [\text{abs}(\hat{\boldsymbol{\beta}}_j)]^{-\gamma}$ , where  $\gamma$  is a positive constant and it usually set to equal one. Eq. (6) also reduces the variance of the conditional mean of the response, and for fixed value of  $n$ , larger values of  $\lambda$  lead to higher bias and less variability in predicted values (Hui, Warton, & Foster, 2014).

### 2.2 RALASSO

The actual variable selection performance of the ALASSO depends on the type of the selected weight. Zou (2006) used the OLS estimator as an initial weights, i.e.,  $w_j = [\text{abs}(\hat{\boldsymbol{\beta}}_{OLS})]^{-\gamma}$ . In the same manner, when other GLMs are used, the initial weight be  $w_j = [\text{abs}(\hat{\boldsymbol{\beta}}_{MLE})]^{-\gamma}$ . These types of weights will not correct in use when there is multicollinearity. From this point, the ratio of the standard error of the ML estimator to the ML estimator was proposed as an initial weight in adaptive LASSO. The advantage of using standard error of the ML  $s_{\hat{\boldsymbol{\beta}}_{ML}}$  is to adjust the adaptive LASSO  $\hat{\boldsymbol{\beta}}$  using ML estimates as an initial value.

Let  $\hat{\boldsymbol{\beta}}_{ML} = (\hat{\beta}_{1(ML)}, \dots, \hat{\beta}_{k(ML)})^T$  be the vector of ML estimate,  $\mathbf{s}_{\hat{\boldsymbol{\beta}}_{ML}} = (s_{1(\hat{\boldsymbol{\beta}}_{ML})}, \dots, s_{k(\hat{\boldsymbol{\beta}}_{ML})})^T$  be the vector of the standard error of the ML, and  $\mathbf{w}_{Ratio} = (w_{1(ratio)}, \dots, w_{k(ratio)})^T$  be the ratio weight vector, where  $w_j = [s_{j(\hat{\boldsymbol{\beta}}_{ML})} / \text{abs}(\hat{\boldsymbol{\beta}}_{j(ML)})]^{-\gamma}$ ,  $j = 1, 2, \dots, k$ . For simplicity, we set  $\gamma = 1$ . Then a coordinate descent method can be used to solve RALASSO. The computation details are given in algorithm 1.

Algorithm 1: The coordinate descent method for the RALASSO.

- (a) Find  $\mathbf{w}_{Ratio}$
- (b) Define  $\mathbf{x}_j^{**} = \mathbf{x}_j / \mathbf{w}_{Ratio}$ ,  $j = 1, 2, \dots, k$ .
- (c) Solve the LASSO for all  $\lambda$  values,

$$\hat{\boldsymbol{\beta}}_{RALASSO}^{**} = \arg \min_{\boldsymbol{\beta}} \left\{ -\sum_{i=1}^n \{y_i \mathbf{x}_i^T \boldsymbol{\beta} - \exp(\mathbf{x}_i^T \boldsymbol{\beta})\} + \lambda \sum_{j=1}^k \mathbf{w}_{j(Ratio)} |\boldsymbol{\beta}_j| \right\}.$$

- (d) Output  $\hat{\boldsymbol{\beta}}_{j(RALASSO)}^* = \hat{\boldsymbol{\beta}}_j^{**} / \mathbf{w}_{Ratio}$ .

### 2.3 Simulation Study

In this section, simulation studies are used to investigate the performance of the RALASSO. Furthermore, we compare RALASSO with LASSO and ALASSO. In all simulations the response variable was generated from Poisson distribution with conditional mean  $\mathbf{x}_i^T \boldsymbol{\beta}$ . All simulation scenarios are replicate 100 times. For every simulation scenario and in each replication we generate training, validation, and testing data. The training data were used for model fitting. The validation data were used to determine the tuning parameters. The testing data were used to evaluate the penalization methods. For each scenario, the observation numbers of the corresponding data sets are denoted by training/validation/testing. Based on the simulated data, we used four metrics to evaluate all penalization methods which were studied in this paper, mean-squared errors for both training data set ( $MSE_{train}$ ) and testing data set ( $MSE_{test}$ ), hits which stands for the number of correctly identified true variables, and false positive (FP) which denotes to the number of zero variables which are wrongly considered as true variables.

Since we investigate a penalization method with both variable selection and multicollinearity, we use three

simulations with different values of the correlation and different numbers of training, validation, and testing observations. In addition, we set different number of variables.

Simulation 1: In this setting, we generate data sets with sample sizes 100/100/400 and 20 explanatory variables. The true parameter vector is specified by  $\beta = (0.5, -0.5, 0.5, -0.6, 0.5, \underbrace{0, \dots, 0}_5, 0.5, -0.5, 0.5, -0.6, 0.5, \underbrace{0, \dots, 0}_5)^T$ . The pairwise correlation was set to  $\rho(x_i, x_j) = 0.8^{|i-j|}$ .

Simulation 2: with  $\beta = (1.25, -0.95, 0.90, -1.10, 0.60, \underbrace{0, \dots, 0}_{15})^T$  and  $\rho(x_i, x_j) = 0.5$  for all values except the

diagonal elements which is equal to 1, we generate data sets with sample sizes 200/200/400 and 20 explanatory variables.

Simulation 3: in this simulation, the true parameter vector is specified by  $\beta = (\underbrace{0.17, \dots, 0.17}_8)^T$ . The pairwise

correlation was set to  $\rho(x_i, x_j) = 0.7^{|i-j|}$ . The generated sample sizes were 1000/1000/1000.

### 3. Results

#### 3.1 Simulation Results

To examine the performance of the RALASSO penalty we compare it with two well-known penalized methods; LASSO and ALASSO. The  $MSE_{train}$  and  $MSE_{test}$  are computed as the criterion of evaluation.

Figures 1 – 2 display the corresponding boxplots of the  $MSE_{train}$  and  $MSE_{test}$  for the three used methods for three simulations, respectively. It is clearly seen that RALASSO has less variability comparing with LASSO and ALASSO.

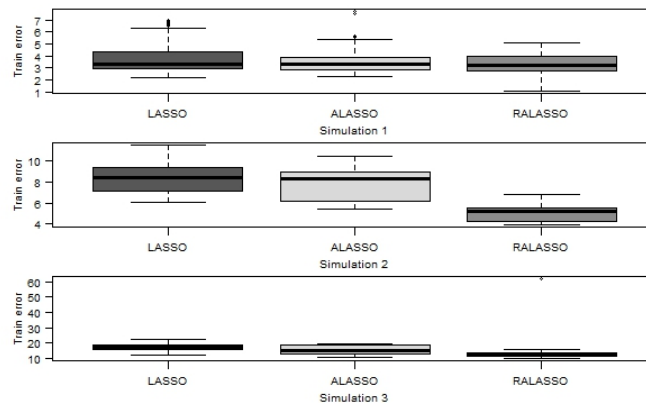


Figure 1. Comparison of median mean-squared error of the training data for three methods

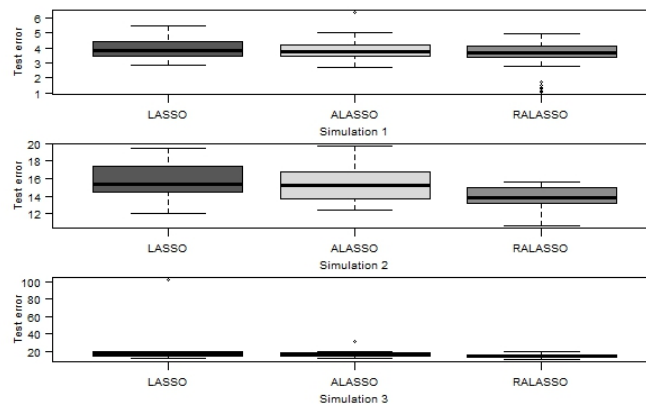


Figure 2. Comparison of median mean-squared error of the testing data for three methods

Table 1 summarizes the median MSE, the median number of hits, and FP. The bold font indicates the best method on MSE, hits, and FP. Table 1 reveals that the RALASSO method produces considerably smaller median  $MSE_{train}$  and  $MSE_{test}$  among all methods in all simulation scenarios. For example, in simulation 1 the median  $MSE_{train}$  of RALASSO is 3.245 which is smaller than 3.314 and 3.346 for ALASSO and LASSO methods,

respectively. Furthermore, the reduction of  $MSE_{test}$  is usually substantial compared to ALASSO. For example, the reduction in simulation 1, simulation 2, and simulation 3 is 2.701%, 9.230%, and 15.062% respectively. Moreover, RALASSO performs well in terms of both  $MSE_{train}$  and  $MSE_{test}$  when the multicollinearity was presented. Besides, from the simulation results we can observe that LASSO came the last method. This is due to its limitation when there is grouping effects between variables.

Table 1. Comparison among methods over 100 replications

	$MSE_{train}$	$MSE_{test}$	Hits	False Positive
Simulation 1				
LASSO	3.346	3.806	6	6
ALASSO	3.314	3.776	7	6
RALASSO	<b>3.245</b>	<b>3.674</b>	<b>10</b>	<b>5</b>
Simulation 2				
LASSO	8.480	15.393	4	7
ALASSO	8.358	15.232	5	7
RALASSO	<b>5.207</b>	<b>13.826</b>	<b>5</b>	<b>6</b>
Simulation 3				
LASSO	17.539	17.533	7	0
ALASSO	14.972	16.531	7	0
RALASSO	<b>12.279</b>	<b>14.041</b>	<b>8</b>	<b>0</b>

For variable selection accuracy, the penalization methods should include all important variables (non-zero variables), hits and FP were used to measure the performance of RALASSO, ALASSO, and LASSO in term of selecting the non-zero variables. From Table 1, RALASSO succeeds in selecting the true non-zero variables in most of the cases in term of hits. For example, in simulation 1 RALASSO selects the all ten non-zero variables. Moreover, when the correlation coefficient varies from small, medium, to high correlation both LASSO and ALASSO select less non-zero variables comparing to RALASSO. We can expect such a result because LASSO and ALASSO have their limitation with grouping effects. In term of FP, RALASSO method usually selects less ineffective variables than ALASSO and LASSO in most cases. To this end, it is obvious from our simulation results that the RALASSO method performs better in term of  $MSE_{train}$  and  $MSE_{test}$  by obtaining smaller values, hits, and FP followed by ALASSO and LASSO for small, medium, and high correlation, and has greater advantage of variable selection with multicollinearity in Poisson regression model.

### 3.2 Real Data Set Application

The real data set which belong to the study of the distribution of freshwater mussels was taken from Sepkoski and Rex (1974). The study aims at the estimation of the numbers of species of mussels in 41 rivers in US by various explanatory variables. The nine explanatory variables are: area, number of stepping stones (intermediate rivers) to 4 major species-source river systems (Alabama-Coosa (AC), Apalachicola (AP), St. Lawrence (SL), and Savannah (SV)), nitrate concentration, hydronium concentration ( $10^{-(pH)}$ ), and solid residue.

In order to investigate the performances of the RALASSO method the data set has been split 50 times at random into a training set of 28 observations and a test set of 12 observations. Model fitting and tuning parameter selection has been done by fivefold cross validation on the training data set. Table 2 shows the median number of explanatory variables selected by each of the LASSO, ALASSO, and RALASSO in the training data set, and the corresponding median  $MSE_{test}$ . It can be seen that RALASSO performs best in term of prediction error where the  $MSE_{test}$  of the RALASSO is approximately 7.45% lower than ALASSO and 9.43% lower than LASSO. Moreover, RALASSO selects less explanatory variables than the other two methods.

Table 2. Comparison among methods for the real data set

	MSE <sub>test</sub>	No. of selected variables
LASSO	10.981	6
ALASSO	10.745	6
RALASSO	<b>9.945</b>	<b>5</b>

#### 4. Conclusion

A study of an improving adaptive LASSO was proposed by applying on Poisson regression model. RALASSO and two penalized methods including LASSO and ALASSO in were compared by using both simulation studies and real data analysis. The simulation and real data results show that the RALASSO is outperforming the other two methods in term of mean-squared error of training and testing data sets and variable selection accuracy. We can conclude that RALASSO more reliable than ALASSO when there is multicollinearity between variables in applying penalized Poisson regression model.

#### References

- Algama, Z. Y. (2012). Diagnostic in poisson regression models. *Electronic Journal of Applied Statistical Analysis*, 5(2), 178-186.
- Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, 32(2), 407-499. <http://dx.doi.org/10.1214/009053604000000067>
- Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456), 1348-1360.
- Fan, Y., & Tang, C. Y. (2013). Tuning parameter selection in high dimensional penalized likelihood. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(3), 531-552. <http://dx.doi.org/10.1111/rssb.12001>
- Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1), 1-22.
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55-67.
- Hossain, S., & Ahmed, E. (2012). Shrinkage and penalty estimators of a Poisson regression model. *Australian & New Zealand Journal of Statistics*, 54(3), 359-373. <http://dx.doi.org/10.1111/j.1467-842X.2012.00679.x>
- Hui, F., Warton, D., & Foster, S. (2014). Tuning parameter selection for the adaptive lasso using ERIC. *Journal of the American Statistical Association*, 00-00. <http://dx.doi.org/10.1080/01621459.2014.951444>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning*. New York: Springer.
- Jianqing, F., & Jinchi, L. (2011). Nonconcave penalized likelihood with NP-dimensionality. *Information Theory, IEEE Transactions on*, 57(8), 5467-5484. <http://dx.doi.org/10.1109/TIT.2011.2158486>
- Lian, H. (2012). Variable selection in high-dimensional partly linear additive models. *Journal of Nonparametric Statistics*, 24(4), 825-839. <http://dx.doi.org/10.1080/10485252.2012.701300>
- Lu, W., Goldberg, Y., & Fine, J. P. (2012). On the robustness of the adaptive lasso to model misspecification. *Biometrika*, 9(3), 717-731. <http://dx.doi.org/10.1093/biomet/ass027>
- Park, M. Y., & Hastie, T. (2007). L1-regularization path algorithm for generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(4), 659-677. <http://dx.doi.org/10.1111/j.1467-9868.2007.00607.x>
- Pourahmadi, M. (2013). *High-Dimensional Covariance Estimation: With High-Dimensional Data*. New Jersey: John Wiley & Sons.
- Qian, W., & Yang, Y. (2013). Model selection via standard error adjusted adaptive lasso. *Annals of the Institute of Statistical Mathematics*, 65(2), 295-318. <http://dx.doi.org/10.1007/s10463-012-0370-0>
- Sampson, J. N., Chatterjee, N., Carroll, R. J., & Müller, S. (2013). Controlling the local false discovery rate in the adaptive Lasso. *Biostatistics*, 14(4), 653-666. <http://dx.doi.org/10.1093/biostatistics/kxt008>
- Sepkoski, J., & Rex, M. (1974). Distribution of Freshwater Mussels: Coastal Rivers as Biogeographic Islands.

*Systematic Zoology*, 23(2), 165-188.

- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), 267-288. <http://dx.doi.org/10.2307/2346178>
- Wang, S., Nan, B., Rosset, S., & Zhu, J. (2011). Random lasso. *The Annals of Applied Statistics*, 5(1), 468-485. <http://dx.doi.org/10.1214/10-AOAS377>
- Wang, Z., Ma, S., Zappitelli, M., Parikh, C., Wang, C. Y., & Devarajan, P. (2014). Penalized count data regression with application to hospital stay after pediatric cardiac surgery. *Statistical Methods in Medical Research*. <http://dx.doi.org/10.1177/0962280214530608>
- Zeng, P., Wei, Y., Zhao, Y., Liu, J., Liu, L., Zhang, R., ... Chen, F. (2013). Variable selection approach for zero-inflated count data via adaptive lasso. *Journal of Applied Statistics*, 41(4), 879-894. <http://dx.doi.org/10.1080/02664763.2013.858672>
- Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476), 1418-1429. <http://dx.doi.org/10.1198/016214506000000735>
- 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. <http://dx.doi.org/10.1111/j.1467-9868.2005.00503.x>

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).