International Journal of Data Science and Analysis
2019; 5(5): 99-103
http://www.sciencepublishinggroup.com/j/ijdsa
doi: 10.11648/j.ijdsa.20190505.14
ISSN: 2575-1883 (Print); ISSN: 2575-1891 (Online)

Penalized Poisson Regression Model Using Elastic Net and Least Absolute Shrinkage and Selection Operator (Lasso) Penality

Josephine Mwikali, Samuel Mwalili, Anthony Wanjoya

Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

Email address:
mwikaliokeri@gmail.com (J. Mwikali), samuel.mwalili@gmail.com (S. Mwalili), awanjoya@gmail.com (A. Wanjoya)

To cite this article:
Josephine Mwikali, Samuel Mwalili, Anthony Wanjoya. Penalized Poisson Regression Model Using Elastic Net and Least Absolute Shrinkage and Selection Operator (Lasso) Penality. *International Journal of Data Science and Analysis.* Vol. 5, No. 5, 2019, pp. 99-103.
doi: 10.11648/j.ijdsa.20190505.14

Received: October 5, 2019; **Accepted**: October 22, 2019; **Published**: October 28, 2019

Abstract: Variable selection in count data using Penalized Poisson regression is one of the challenges in applying Poisson regression model when the explanatory variables are correlated. To tackle both estimate the coefficients and perform variable selection simultaneously, Lasso penalty was successfully applied in Poisson regression. However, Lasso has two major limitations. In the p > n case, the lasso selects at most n variables before it saturates, because of the nature of the convex optimization problem. This seems to be a limiting feature for a variable selection method. Moreover, the lasso is not well-defined unless the bound on the L1-norm of the coefficients is smaller than a certain value. If there were a group of variables among which the pairwise correlations are very high, then the lasso tends to select only one variable from the group and does not care which one is selected. To address these issues, we propose the elastic net method between explanatory variables and to provide the consistency of the variable selection simultaneously. Real world data and a simulation study show that the elastic net often outperforms the lasso, while enjoying a similar sparsity of representation. In addition, the elastic net encourages a grouping effect, where strongly correlated predictors tend to be in the model together.

Keywords: Penalized, Poisson Regression, Elastic Net Penalty, Lasso

1. Introduction

In modern data analysis problem, we had number of parameters greater than number of observation leading to high dimensional problems.

Health, finance, economics and sports to mention a few were some of the areas that had benefited drastically from the ever increasing level of technology. This has seen an enormous amount of data derived with two dimensions the number of both variable and observation.

This is different from the normal dataset we encounter for statistical analysis having many observations on a few variables. This type of dataset, however, comes with new challenges because of its complexity and cannot simply apply classically statistical methods such as Poisson regression, ineffective, because statistical issues associated with modeling high dimensional data include model over fitting, estimation instability, computational difficulty [1].

The criteria for evaluating the quality of a model will differ according to the circumstances. Typically, the following two aspects are important:

Accuracy of prediction on future data- it is hard to defend a model that predicts poorly.

Interpretation of the model- scientists prefer a simpler model because it explains more light on the relationship between response and covariates. Parsimony is especially an important issue when the number of predictors is large [2].

Ordinary Least squares does poorly in both prediction and interpretation that lead to introduction of Penalized techniques to improve OLS. 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 [3]. To do this, various penalized methods have been proposed. The least absolute shrinkage and selection operator (Lasso) to estimate the regression coefficients through L2- norm penalty [4].

Lasso has advantages in that it can provide a very good prediction accuracy, because shrinking and removing the coefficients can reduce variance without a substantial increase of the bias, this is especially useful when you have a small number of observation and a large number of features [5].

In terms of the tuning parameter λ we know that bias increases and variance decreases when λ increases, indeed a trade-off between bias and variance has to be found. Moreover, the Lasso helps to increase the model interpretability by eliminating irrelevant variables that are not associated with the response variable, this way also overfitting is reduced. This is the point where we are more interested in because in this paper the focus is on the feature selection task [6].

The elastic net penalty which is based on a combined penalty of Lasso and ridge regression penalties in order to overcome the drawbacks of using the Lasso on its own [7].

Usually, in high dimensional data the explanatory variables are correlated. If there is a group of highly correlated variables, the lasso will randomly select only one variable from this group and drop the rest whereas elastic net will select the whole group of the highly correlated explanatory variables [8].

In this paper we propose a new regularization technique which we call the elastic net.

Similar to the lasso, the elastic net simultaneously does automatic data reduction and continuous shrinkage, and is able to select groups of correlated variables [9].

The remainder of the paper is structured as follows: Section 2 discusses the existing

literature, Section 3 presents an overview of the Penalized Poisson regression models and outlines the lasso and Elastic regression methods. Section 4 describes the simulation study and real data analysis. Section 5 results obtained from the real data and simulated data. Section 6 concludes the study.

2. Literature Review

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 [10, 11].

A procedure called penalization, which is always used in variables selection in high dimensional data, attaches a penalty term $p_\lambda(B)$ the log-likelihood function to get a better estimate of the prediction error by avoid over fitting. Recently, there is growing interest in applying the penalization method in the Poisson regression models. An efficient algorithm for the estimation of a generalized linear model including Poisson regression with a convex penalty [12]. Stein-type shrinkage estimator for the parameters of Poisson regression model [13]. A combination of minimax concave and ridge penalties and a combination of smoothly clipped absolute deviation and ridge penalties [14]. 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.

Equation 1

$$f(y_i) = \frac{e^{-\theta_i}\theta_i{}^{y_i}}{y_i!}$$

and the conditional mean parameter $\theta_i = \exp(x_i'\beta)$. Under the assumption of independent observation, the log-likelihood function is given by

Equation 2

$$\sum_{i=1}^{n} y_i x_i'\beta - \exp(x_i'\beta) - lny_i!.$$

3. Methodology

3.1. Penalized Poisson Regression Model

The penalized Poisson regression is defined by;

Equation 3

$$PPR = l(\beta) + \lambda p(\beta)$$

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

Since the value of λ is depends on the data, it can be calculated using cross- validation method [9].

Before solving the PPR, it is worth to make standardization to x_i so that

$$\frac{1}{n}\sum_{i=1}^{n} x_{ij} = 0 \text{ and } \sum_{i=1}^{n} x_{ij}^2 \text{ for j =1, 2,…k}$$

This makes intercept β_0 equal to zero.

3.2. Lasso Regression

The lasso (least absolute shrinkage and selection operator) is a regression analysis method that performs both variable selection and regularization, in order to enhance the prediction accuracy and interpretability of the regression model by altering the model fitting process to select only a subset of the provided covariates for use in the final model rather than using all of them.

The Lasso for the Poisson regression model was originally proposed [15].

Lasso is able to achieve both of these goals by forcing the sum of the absolute value of the regression coefficient to be less than a fixed value which 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. Cross-validation methods can be used for identifying which of these two

techniques is better on a particular data set.

The method applies a shrinking (regularization) process where it penalizes the coefficients of the regression variables shrinking some of them to zero. During features selection process the variables that still have a non-zero coefficient after the shrinking process are selected to be part of the model. The goal of this process is to minimize the prediction error.

The lasso estimator is

Equation 4

$$\beta_{lasso} = argMIn_{\beta}(l(\beta) + \lambda \sum_{j=1}^{k} |\beta_j|)$$

where $\lambda \geq 0$ is the tuning parameter that controls the strength of the penalty assumes great importance.

For large values of λ coefficients are forced to be exactly equal to zero. This way, dimensionality can be reduced. The larger the parameter λ the more the number of coefficients are shrunk to zero.

3.3. Elastic-net

The Elastic net is a regularized regression method that linearly combines the L1 and L2 penalties of Lasso and Ridge methods.

The elastic net overcomes the limitation of Lasso selects at most n variables before it saturates. Also if the group is highly correlated variables then the Lasso tend to select one variable from the group and ignore the others. The elastic net estimator which is a combination between the ridge and the lasso penalty [12].

The second term (ridge penalty) encourages highly correlated variables to be averaged, while the first term (the LASSO penalty) encourages a sparse solution in the coefficients of these average variables. The elastic net estimator for Poisson regression model is elastic net estimator is depended on non-negative two tuning parameters λ_1, λ_2 and leads to penalized Poisson regression solution.

Equation 5

$$\beta_{elastic} = argMIn_{\beta}(-l(\beta) + \lambda_1 \sum_{j=1}^{k} |\beta_j| + \lambda_2 \sum_{j=1}^{k} |\beta_j|^2)$$

3.4. Model Testing

The developed model will be tested on real count data set of choice to determine its performance compared to its linear counterpart. The statistics MSE will be used to make a conclusion on the performance of the models. The model with least mean-squared errors for the test data (MSE) will be considered the best. The following test statistic was used as performance measurements between the two models.

Equation 6

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{Z}_i)$$

4. Data Description

To examine the prediction accuracy and variable selection of the elastic net we compare it with the Lasso penalties on Penalized Poisson regression using simulated data and real data.

In simulations the response variable were generated from Poisson distribution with conditional mean θ_i. Our simulated data consist of a training set, validation set and testing set.

The training data were used for model fitting; the validation data were used to determine

the tuning parameters and the testing data were used to evaluate the penalization methods.

The observation numbers of the corresponding data sets were denoted by training/validation/testing.

Also the MSE was computed on test data. 50 datasets were simulated consisting of 50/50/400 observation.

The real data set were from a study of prostate cancer [16]. The predictors are eight clinical measures: log cancer volume (lcavol), log prostate weight (lweight), age, log of the amount of benign prostatic hyperplasia (lbph), seminal vesicle invasion (svi), log capsular penetration (lcp), Gleason score (gleason) and percentage Gleason score 4 or 5 (pgg 45). The response is the log of prostate specific antigen (lpsa).

5. Empirical Result

Elastic net is more accurate than the lasso where grouped selection is required, the elastic net behaves like the oracle. The additional grouped selection ability makes the elastic net a better variable selection method than the lasso. In each case, Table 1 reveals that the elastic method produces considerably smaller median MSE and standard deviation among all methods in all cases.

Table 1. *Comparison of MSE Median and their Standard deviation among two methods.*

Statistic	MSE	Std
Lasso	46.6	3.96
Elastic net	34.5	1.64

It is obvious from our simulation results that the elastic net performs better in term of MSE by obtaining smaller values high correlation. Elastic net has greater advantage of variable selection with grouping effects in Poisson regression model.

The lasso and elastic net were all applied to the prostate cancer data, in order to enable a fair comparison, typically, the dataset was randomly partitioned into a training dataset, which comprised 70% of the samples, and a test dataset, which consisted of 30% of the samples.

In order to get the best value of the pair (λ-1, λ_2), 10-fold cross validation were carried out on the training data. We then compared the performance of those methods by computing their prediction mean squared error on the test data. All the applications were conducted in R using the glmnet package. Table 2 shows the median number of explanatory variables selected by each of the elastic net and lasso in the test data set, and the corresponding median MSE.

It can be seen that Elastic net performs best in term of prediction error where the MSE of the elastic net is approximately 0.21% lower than Lasso. Moreover, elastic net selects less explanatory variables than the other method

Table 2. *Comparison of Median MSE for the Regression methods.*

MSE	Min	1st Qu	Median	Mean	3rd Qu	Max
Lasso	0.6056	0.6390	0.7019	0.7257	0.7770	0.9570
Elastic net	0.6113	0.6310	0.6991	0.7262	0.7835	0.9649.

The lasso includes lcavol, lweight lbph, lcp, svi, and pgg45 in the final model, while the elastic net selects lcavol, lweight, svi, lcp, and pgg 45.

The prediction error of the elastic net is about 10 percent lower than that of the lasso.

Table 3. *Variable selection among the methods for Real data.*

Variables	Lasso	Elastic
(Intercept)	-0.359568455	-0.313754493
lcavol	0.464825516	0.433777648
lweight	0.557305441	0.550976927
Age		
lbph	0.014129093	
svi	0.333693029	0.388507425
lcp		
gleason		
Pgg 45	0.005352502	0.005773321

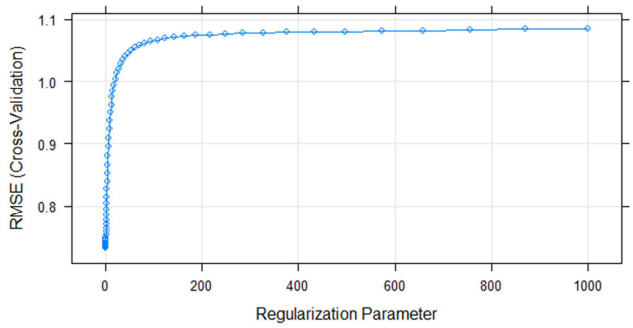

Figure 1. *Lasso.*

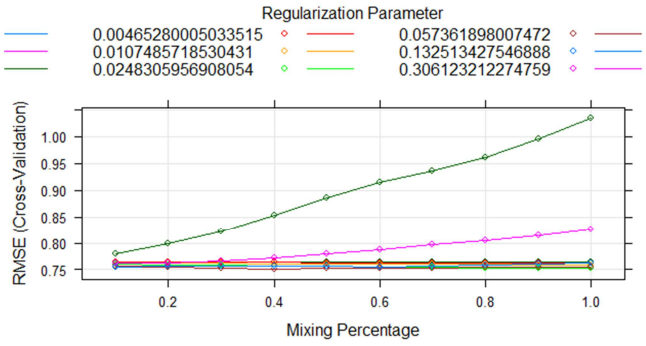

Figure 2. *Elastic net.*

6. Conclusion

A study of elastic net was proposed by applying on Penalized Poisson regression model. Elastic net and Lasso were compared by using simulation studies and real data application.

Both the simulation and real data results show that the Elastic produces a model with good prediction accuracy and also outperforming the Lasso in term of MSE of test data and variable selection accuracy.

Elastic net encourages a grouping effect. We can conclude that Elastic net combines feature elimination from lasso and feature coefficient reduction from the ridge model to improve your model's prediction. Elastic net can be used as a generalization of Lasso which has been shown to be of great importance for model fitting and variable selection in high dimension data.

References

[1] M. Pourahmadi, High-dimensional covariance estimation: with high-dimensional data, vol. 882. John Wiley & Sons, 2013.

[2] S. Hossain and E. Ahmed, "Shrinkage and penalty estimators of a Poisson regression model," Aust. N. Z. J. Stat., vol. 54, no. 3, pp. 359–373, 2012.

[3] M. El Anbari and A. Mkhadri, "Penalized regression combining the L 1 norm and a correlation based penalty," Sankhya B, vol. 76, no. 1, pp. 82–102, 2014.

[4] H. D. Bondell and B. J. Reich, "Simultaneous regression shrinkage, variable selection, and supervised clustering of predictors with OSCAR," Biometrics, vol. 64, no. 1, pp. 115–123, 2008.

[5] R. Tibshirani, "Regression shrinkage and selection via the lasso," J. R. Stat. Soc. Ser. B, vol. 58, no. 1, pp. 267–288, 1996.

[6] O. Troyanskaya et al., "Missing value estimation methods for DNA microarrays," Bioinformatics, vol. 17, no. 6, pp. 520–525, 2001.

[7] Z. Wang, S. Ma, and C.-Y. Wang, "Variable selection for zero-inflated and overdispersed data with application to health care demand in Germany," Biometrical J., vol. 57, no. 5, pp. 867–884, 2015.

[8] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," J. R. Stat. Soc. Ser. B (statistical Methodol., vol. 67, no. 2, pp. 301–320, 2005.

[9] Y. Fan and C. Y. Tang, "Tuning parameter selection in high dimensional penalized likelihood," J. R. Stat. Soc. Ser. B (Statistical Methodol., vol. 75, no. 3, pp. 531–552, 2013.

[10] Z. Y. Algamal, "Diagnostic in poisson regression models," Electron. J. Appl. Stat. Anal., vol. 5, no. 2, pp. 178–186, 2012.

[11] Z. Y. Algamal and M. H. Lee, "Penalized logistic regression with the adaptive LASSO for gene selection in high-dimensional cancer classification," Expert Syst. Appl., vol. 42, no. 23, pp. 9326–9332, 2015.

[12] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," J. Stat. Softw., vol. 33, no. 1, p. 1, 2010.

[13] K. Hoffmann, "Stein estimation—a review," Stat. Pap., vol. 41, no. 2, p. 127, 2000.

[14] F. Xue and A. Qu, "Variable Selection for Highly Correlated Predictors," arXiv Prepr. arXiv1709.04840, 2017.

[15] M. Y. Park and T. Hastie, "L1-regularization path algorithm for generalized linear models," J. R. Stat. Soc. Ser. B (Statistical Methodol., vol. 69, no. 4, pp. 659–677, 2007.

[16] G. N. Collins, R. J. Lee, G. B. McKelvie, A. C. N. Rogers, and M. Hehir, "Relationship between prostate specific antigen, prostate volume and age in the benign prostate," Br. J. Urol., vol. 71, no. 4, pp. 445–450, 1993.