A STUDY ON COMPARISON AMONG RIDGE, LASSO, AND ELASTIC NET REGRESSIONS

B.Sarojamma[1]
[1]Assistant Professor,
Department of Statistics, S.V.University, Tirupati-517502, A.P.

K. AnilKumar[2].
[2]Teaching Professional,
National Institute of Agricultural Extension Management,
MANAGE, Rajendranagar, Hyderabad- 500030. T.S.

Abstract: If there are two or more independent variables are explaining the pragmatic behavior of either one or more dependent variables then Regression analysis is the best option for fitting a model as well as for forecasting. There are so many types of regression techniques in literature, some of the popular Regression techniques are Linear Regression, Logistic Regression, Polynomial Regression, Stepwise Regression, Ridge regression, Lasso Regression, Elastic Net Regression etc.

In the present study we are comparing Ridge regression, lasso Regression and Elastic Net Regression using Durbin Watson test and empirically tested by using real data.

Keywords: Linear Regression, Multiple Regression, Ridge regression, Lasso Regression, Elastic Net Regression.

1. INTRODUCTION:

If we continue to draw from OLS (Ordinary Least Square) as our only approach to linear regression technique methodologically speaking, we are still within the late 1800's and 1900's time frame. With advancements in computing technology, regression techniques can be used in a wide variety of different statistical techniques which has led to development of new tools and techniques. In current data analysis, we usually find data with enormous number of independent variables and are need better regression techniques to handle this high-dimensional modelling.

2. REVIEW OF LINEAR REGRESSION ANALYSIS:

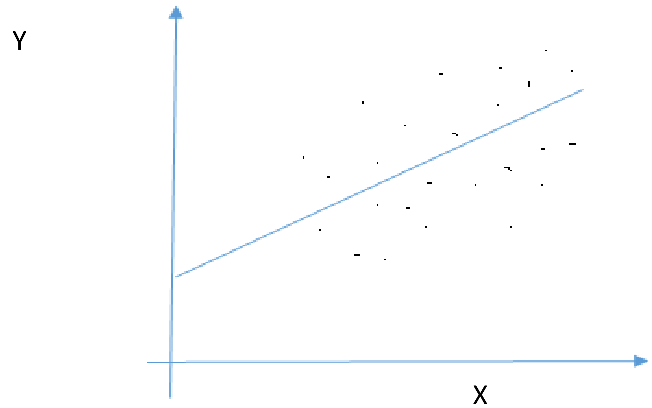
Simple linear regression: a model with single independent variable is called simple linear regression, the model of the form as$Y = \beta_0 + \beta_1 X + \epsilon$. The term β_0 represents the Y's intercept value, the coefficient β_1 denotes the slope of the line or termed as regression coefficient, the term X is independent variable and ϵ is an error term. The error term is the value, which needs to correct for a prediction error between the observed and predicted value.

Multiple linear regression: A multiple linear regression model is essentially same as a simple linear regression except that there can be multiple coefficients and independent variables. The model of the form

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \ldots + \epsilon.$$

The interpretation of the coefficient terms are slightly differs from simple linear regression model.

OLE (Ordinary Least Squares Estimation): the method is used for estimating the unknown parameters involved in a linear regression model. The goal of OLS is to minimize the difference between observed responses in some arbitrary dataset and the responses predicted by the linear approximation of the data. $Y_n = \sum_{i=0}^{k} \beta_i X_{ni} + \epsilon_n$

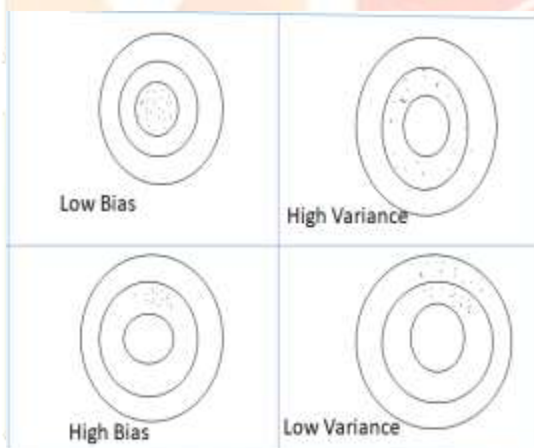


By visualization, one can say that the sum of the vertical distances between each data point in the dataset and the corresponding point on the regression line. The smaller the difference, the model fits better for the given dataset.

Understanding the error:

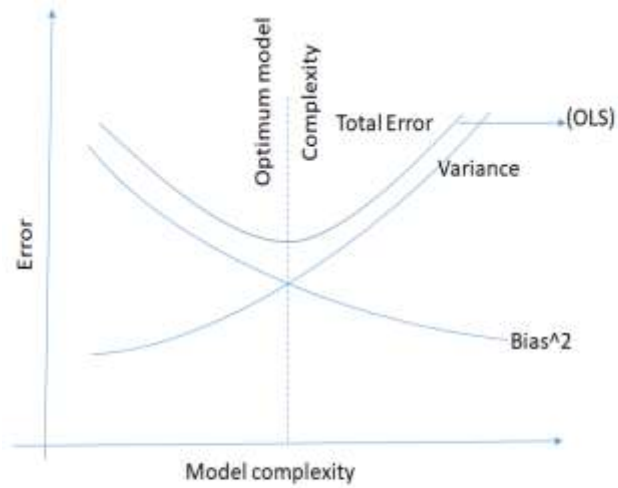
The sum of squares are a representation of the error for our OLS regression model. Usually prediction errors in linear regression models can be decomposed into two main subcomponents, are error due to "bias" and error due to "variance". In general one can take care about overall error but not specific decomposition of error. Understanding how the difference sources of error leads to bias and variance in prediction error helps us in better model fitting with more accuracy in predictive modelling. For this one can use techniques like ridge, Lasso and Elastic net regression etc.

Bias & Variance Trade off:



Error due to bias: The error due to bias is taken as the difference between the expected prediction of our model and the corrected value which we are trying to predict.

Error due to variance: the error due to variance is taken as the variability of model prediction for a given data point.



From the above figure, with minimum error the model complexity get increased along with bias square. On the other hand error and variance both are in same direction with respect to model complexity. Hence by understanding these two types of errors can help us in diagnose model results and avoid the problem of under or over fitting of the model.

Ridge Regression: Gauss Markov theorem states that OLS estimates are unbiased with the smallest mean square error. From this it arise question is there a biased estimator with smaller mean square error for the good model fit? To answer this question shrinkage come into the picture.

Let us replace OLS estimates β_k with some slight change as $\beta_k^1 = \frac{1}{1+\lambda}\beta_k$; if $\lambda = 0$, we will get the OLS estimates back. If λ gets really large, the parameter estimate approaches a minimum value (zero). Here λ is referred to as the shrinkage estimator (ridge constant).

Ridge Regression minimizes that in constrained form as

$$\sum_{i=1}^{N}\left(y_i - \sum_j \beta_j x_{ij}\right)^2 \text{subject to } \sum_j \left\|\beta_j\ \right\|^2 \leq t$$

The above equation takes in matrix form $RSS(\lambda) = (Y - X\beta)^T(Y - X\beta) + \lambda\beta^T\beta$; the ridge estimate of β is given by $\widehat{\beta^{ridge}} = (X^TX + \lambda I)^{-1}X^TY$.

LASSO Regression (Least Absolute selection and shrinkage operator): The LASSO combines some of the shrinking advantages of ridge regression with variable selection. The difference between the LASSO and the Ridge regression is that ridge uses $\|\beta\|^2$ penalty whereas the LASSO uses $\|\beta\|$ penalty, even though these l_1 and l_2 looks similar, the solution behaves quite different.

The lasso estimate is defined by

$$\hat{\beta}^{lasso} = argmin \sum_{i=1}^{N}\left(y_i - \sum_j \beta_j x_{ij}\right)^2 \qquad \text{subject to} \sum_j \left\|\beta_j\right\| \leq t.$$

Elastic net regression:

The elastic net estimator $\hat{\beta}$is minimized of the form

$$\hat{\beta}^{elastic\ net} = \arg min \sum_{i=1}^{N}\left(y_i - \sum_j \beta_j x_{ij}\right)^2 + \lambda_2\left\|\beta_j\right\|_2 + \lambda_1\left\|\beta_j\right\|_1 \qquad \ldots\ldots\ldots\ldots\ldots (2.1)$$

The procedure is viewed as a penalized least square method. Let $\alpha = {\lambda_2}/{(\lambda_1 + \lambda_2)}$

The optimization problem equivalent to equation (2.1) can be written as

$$\hat{\beta}^{elastic\ net} = \arg min \sum_{i=1}^{N}\left(y_i - \sum_j \beta_j x_{ij}\right)^2, \text{ subject to} (1-\alpha)\|\beta\|_1 + \alpha\|\beta\|_2 \leq t.$$

3. EMPIRICAL INVESTIGATION:

In present study we are fitted Ridge Regression, Lasso Regression and Elastic net Regression models for intraday data of stock opening from 1st January 2016 to 1st December 2017. When the data having multicollinearity, one can use ridge regression to overcome this problem by shrinkage parameter λ. the model for ridge regression is given by

$argmin\ \sum(y - x\beta)^2 + \lambda\sum\beta_2^2$ Where y is dependent term, x is an independent term, β is regression coefficient and λ is Shrinkage parameter.

The fitted form of the above equation is $argmin\ \sum_{i=1}^{477}\left(y - x\widehat{\beta_\iota}\right)^2 + \lambda\sum\hat{\beta}^2$; where $\widehat{\beta_\iota}$ is estimated value by using the following formula.

$$\widehat{\beta_\iota} = \frac{sum\ of\ consequent\ four\ observation\ of\ x_i}{sum\ of\ consequent\ four\ observation\ of\ y_i}, \qquad i = 1, 2, 3 \ldots..$$

Lasso regression: Least Absolute Shrinkage and Selection Operator (LASSO) Regression, which is similar to ride regression, the equation for LASSO is as given below

$$argmin \sum\left(y_i - X\widehat{\beta_\iota}\right)^2 + \lambda\sum\widehat{\beta_\iota}$$

where y_i is dependent variable, X is independent variable, λ is parameter

The fitted equation for opening sensex intraday data for $\lambda = 0.1$ is as follows

$$argmin \sum_{i=1}^{477}\left(y_i - X\widehat{\beta_\iota}\right)^2 + 0.1\sum\widehat{\beta_\iota}$$

Elastic net regression: It is hybrid form Lasso and ridge regression techniques. The equation for elastic net regression is of the form as

$$\hat{\beta}^{elastic\ net} = argmin\sum\left(y_i - \sum\beta_j x_{ij}\right)^2 + \lambda_2\left\|\beta_j\right\|_2 + \lambda_1\left\|\beta_j\right\|_1$$

where y is dependent term, x is independent term, λ_1, λ_2 are parameters.

The fitted equation of elastic net regression is

$$\hat{\beta}^{elastic\ net} = argmin\sum_{i=1}\left(y_i - \sum\hat{\beta}_j x_{ij}\right)^2 + \lambda_2\left\|\beta_j\right\|_2 + 0.1\left\|\beta_j\right\|_1$$

Here λ_2 taking three values 0.1, 0.5 and 0.9.

Durbin-Watson Statistic: the formula for Durbin-Watson test is

$$dpd = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T}\left(e_{i,t} - e_{i,t-1}\right)^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}e_{i,t}^2}$$

where $e_{i,t}$ and $e_{i,t-1}$ are error terms of t^{th} and $(t-1)^{th}$ terms.

The Durbin-Watson test static values for five models are as follows:

Model	Durbin-Watson test static value
Ridge regression model	0.99425
LASSO regression model	0.86944
Elastic net regression with $\lambda_1 = 0.1, \lambda_2 = 0.1$	1.5369
Elastic net regression with $\lambda_1 = 0.1, \lambda_2 = 0.5$	1.5340
Elastic net regression with $\lambda_1 = 0.1, \lambda_2 = 0.9$	1.5337

From the above table we can say that for elastic regression case Durbin-Watson test value is almost same up to two decimal places for different values ofλ_2.

4. SUMMARY AND CONCLUSIONS:

For stock exchange opening intraday data, we fitted five regression models are as below

$$\beta^{Ridge\ reg} = argmin \sum_{i=1}^{477} \left(y - x\hat{\beta}\right)^2 + 0.1 \sum \hat{\beta}_2^2$$

The above equation is of the form for ridge regression. LASSO regression is given by

$$\hat{\beta}^{LASSO} = argmin \sum_{i=1}^{477} \left(y_i - X\hat{\beta}_\iota\right)^2 + 0.1 \sum \hat{\beta}_\iota$$

The elastic net regression for various values for λ_2 like 0.1, 0.5 and 0.9 respectively are given as below.

$$\hat{\beta}^{elastic\ net} = argmin \sum \left(y_i - \sum \hat{\beta}_J x_{ij}\right)^2 + 0.1\|\hat{\beta}_J\|_2 + 0.1\|\hat{\beta}_J\|_1$$
$$\hat{\beta}^{elastic\ net} = argmin \sum \left(y_i - \sum \hat{\beta}_J x_{ij}\right)^2 + 0.5\|\hat{\beta}_J\|_2 + 0.1\|\hat{\beta}_J\|_1$$
$$\hat{\beta}^{elastic\ net} = argmin \sum \left(y_i - \sum \hat{\beta}_J x_{ij}\right)^2 + 0.9\|\hat{\beta}_J\|_2 + 0.1\|\hat{\beta}_J\|_1$$

For the above models Durbin-Watson test static is calculated, based on this statistic we can choose ridge regression model is best model for the intraday data.

References:

[1]ALAN J. MILLERJ. R. Statist. Soc. A (1984), “Selection of Subsets of Regression Variables”,Part 3, pp. 389-425.
[2] Copas, J. B. (1983), “Regression, prediction and shrinkage”, JR. Statist. Soc. B, 45, 311-354.
[3]Frank, I. and Friedman, J. (1993) A statistical view of some chemo metrics regression tools. Technometrics, 35,109–148.
[4] Hastie, T., Tibshirani, R. and Friedman, J. (2001), “The Elements of Statistical Learning; Data Mining, Inference and Prediction”, New York: Springer.
[5] Joseph O Ogutu ,Torben Schulz-Streeck, “Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions”, BMC Proceedings 2012.
[6] Hoerl, A. and Kennard, R. (1988), “Ridge regression. In Encyclopedia of Statistical Sciences”, vol. 8, pp.129–136. New York: Wiley.
[7] Tibshirani, R. (1996), “Regression shrinkage and selection via the lasso”, J. R. Statist. Soc. B, 58, 267–288.