ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

Regression: An Efficient Data Analysis Technique

Dr. Anuprita Deshmukh[1], Dr. S. D. Gore[2]

Assistant Professor, Sinhgad College of Science, Savitribai Phule Pune University, Pune, Maharashtra, India[1]

Rtd. Professor, Savitribai Phule Pune University, Pune, Maharashtra, India [2]

ABSTRACT: Data Analysis is very important in order to predict the future. Data analyzing task is treditionaly done by manually using statistical methods / tools / techniques. Since it was done by manually there was some error / bugs and it was time consuming. The functions which requires large amount of calculations cannot be done if the fast result / prediction / output is expected. Now because of usage of computer softwares this problem is removed. And hence regression can be done with much more efficiently and it can give results more accurately.

KEYWORDS: Data, Data Analysis, Regression Models / Types

It is easily visible that the huge data is available for analyzing. The main recourse for this collection of data in widely spread internet. This data in public domain can be used for further usage.

I. NEED OF DATA ANALYSIS FOR PREDICTION

Now a day huge data is available in all types of the business / Research / etc. Increasing volumes of data is collected everywhere in the market. This data is available in different types. This data has to be used to produce valuable insights / predictions. The industry / Researchers are interested in the predictive analysis of the data to increase their competitive advantage. Predictions are important in the area such as Detecting fraud, optimizing marketing campaigns reducing risk so on and so forth. It is used in the industries like Banking & Financial Services, Manufacturing, and Health Insurance etc. Prediction analysis is making accurate / approximate prediction about upcoming unidentified future events. Predictive analysis uses many techniques such as artificial intelligence, data mining, statistical methods (e.g. regression).

II. REGRESSION TECHNIQUES

The analysis of the data is depends upon the variables of the data. According to the data various techniques can be used. Statistical method regression is used finding the relation between these variables. Regression analysis is an important tool for modeling and analyzing data. As per the relation of the variables in the data values the mathematical equation is produced with the help of different models of the regression. Then the dependent and independent of the data are visible. The graphs can be drawn using this regression, so that it is easily understood by non-statistical analyst. Depends upon situation models can be applied while performing the analysis.

Regression analysis is a statistical process for estimating the relationships among variables. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables (or 'predictors'). More specifically, regression analysis helps one understand how the typical value of the dependent variable (or 'criterion variable') changes when any one of the independent variables is varied, while the other independent variables are held fixed. Most commonly, regression analysis estimates the conditional expectation of the dependent variable given the independent variables – that is, the average value of the dependent variable when the independent variables are fixed.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

Regression analysis is widely used for prediction and forecasting, where its use has substantial overlap with the field of machine learning. Regression analysis is also used to understand which among the independent variables are related to the dependent variable, and to explore the forms of these relationships. In restricted circumstances, regression analysis can be used to infer causal relationships between the independent and dependent variables.

III. REGRESSION MODELS

There are many models are available in regression. As stated above any one or combination can be used to analysis the data. The models / types are given below. In this paper the types are introduced next paper I will discuss in detail with its applications.

The models / types of the regression:

1. Linear Regression
2. Logistic Regression
3. Polynomial Regression
4. Stepwise Regression
5. Ridge Regression
6. Lasso Regression
7. ElasticNet Regression

3.1 Linear Regression

It analyzes the relationship between the response or dependent variable and a set of independent or predictor variables. This relationship is expressed as an equation that predicts the response variable as a linear function of the parameters. These parameters are adjusted so that a measure of fit is optimized. Much of the effort in model fitting is focused on minimizing the size of the residual, as well as ensuring that it is randomly distributed with respect to the model predictions.[6]

The simple (or bivariate) Linear Regression Model is designed to study the relationship between a pair of variables that appear in a data set. The multiple LRM is designed to study the relationship between one variable and several of other variables.

Using linear regression, it can be predict one variable from the values of other variable. One single independent variable is used to find the value of other dependent variable.

3.2 Logistic Regression

Logistic regression is used to find the probability of event=Success and event=Failure. We should use logistic regression when the dependent variable is binary (0/ 1, True/ False, Yes/ No) in nature. Here the value of Y ranges from 0 to 1 and it can represented by following equation.[6]

If there are more than one independent variable is present in the data then logistic regression is used to predict the unknown value of the other variable.

A regression equation is a polynomial regression equation if the power of independent variable is more than 1. In this regression technique, the best fit line is not a straight line. It is rather a curve that fits into the data points.[6]

3.3 Stepwise Regression

This form of regression is used when we deal with multiple independent variables. In this technique, the selection of independent variables is done with the help of an automatic process, which involves no human intervention. [6] Here the as the name tells the regression is done stepwise.

3.4 Ridge Regression

Ridge Regression is a technique used when the data suffers from multi-collinearity (independent variables are highly correlated). In multi-collinearity, even though the least squares estimate (OLS) are unbiased; their variances are large which deviates the observed value far from the true value. By adding a degree of bias to the regression estimates, ridge regression reduces the standard errors. [6]

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

3.5 Lasso Regression

Similar to Ridge Regression, Lasso (Least Absolute Shrinkage and Selection Operator) also penalizes the absolute size of the regression coefficients. In addition, it is capable of reducing the variability and improving the accuracy of linear regression models. [6] It performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the statistical model it produces.

3.6 ElasticNet Regression

ElasticNet is hybrid of Lasso and Ridge Regression techniques. It is trained with L1 and L2 prior as regularizes. Elastic-net is useful when there are multiple features which are correlated. Lasso is likely to pick one of these at random, while elastic-net is likely to pick both. [6]

IV. CONCLUSION

Regression technique is useful for data analyzing and making the prediction. All the types of the regression can be used as per the situation / requirement.

REFERENCES

[1] https://www.sas.com/en_us/insights/analytics/predictive-analytics.html
[2] https://halobi.com/2016/07/descriptive-predictive-and-prescriptive-analytics-explained/
[3] http://www.predictiveanalyticstoday.com/what-is-predictive-analytics/
[4] https://www.tutorialspoint.com/data_mining/dm_classification_prediction.htm
[5] https://en.wikipedia.org/wiki/Regression_analysis
[6] https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/
[7] Dan Campbell, Sherlock Campbell, Introduction to Regression and Data Analysis
[8] Predictive Analytics: Bringing The Tools To The Data An Oracle White Paper September 2010
[9] Flavia FECHETE, Prof.univ.dr. Anisor Nedelcu, Analysis Of The EconomIc Performance Of A Organization Using Multiple Regression , International Conference Of Scientific Paper, Afases 2014