Journal of The Gujarat Research Society

*ISSN: 0374-8588*
*Volume 22 Issue 1, January 2020*

# An Overview on Several Classification of Regression Algorithm in Machine Learning

***Dr Vipin Solanki***

***SOEIT, Sanskriti University, Mathura, Uttar Pradesh, India***

***Email Id-hodmaths@sanskriti.edu.in***

***ABSTRACT: The algorithms for supervised machine learning include regression algorithms and are part of the algorithm for the learning of machines. The model connections and dependencies between input and output functionality in order to predict new data values are once the main elements of the supervised learning Algorithm. This study examines the fundamental aspects of machine learning, such as machine learning idea or description, followed by an explanation of the many kinds of Machine Learning (ML), such as supervising ML, unsurprising ML and enhancing ML. This review, in conclusion, examines the usage of machine learning algorithms in various applications. This article then discusses the many types of regression algorithms used in ML. In the 21st century ML is the most promising career. They have several excellent possibilities for paid jobs. In addition, there are dramatically altering the automation industry with the possible range of machine learning. There is much promise for machine learning in India as well. Machine learning therefore has a successful life or scope in the ML area and helps to the growth of the digital world.***

***KEYWORDS: Algorithm, Data, Machine Learning (ML), Prediction, Supervised.***

## INTRODUCTION

ML learns the computer programmed which improves both as data results and experience over a period of time. Components of artificial intelligence are taken into consideration. Ml algorithm builds models based on model data and specifies the data to be developed in order for judgments or forecasts to be decided without specification[1]. Machine learning algorithms are being used with emails, medicine and computer vision in a wide range of applications where standard algorithms emerge to achieve an impossible or tough task[2]. The various applications of the machine learning are given in the Figure 1.

Journal of The Gujarat Research Society

ISSN: 0374-8588
Volume 22 Issue 1, January 2020

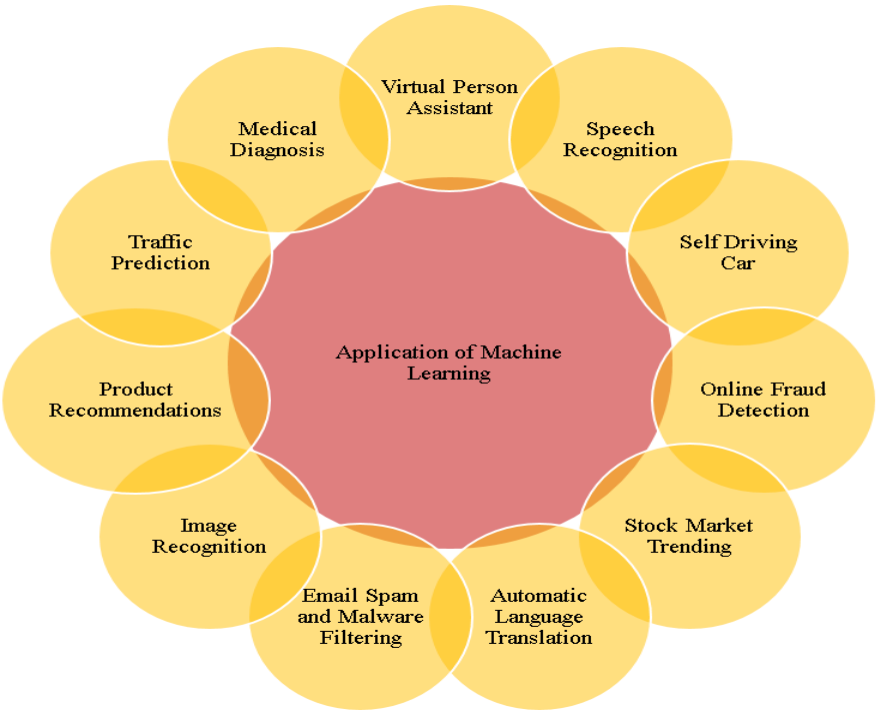
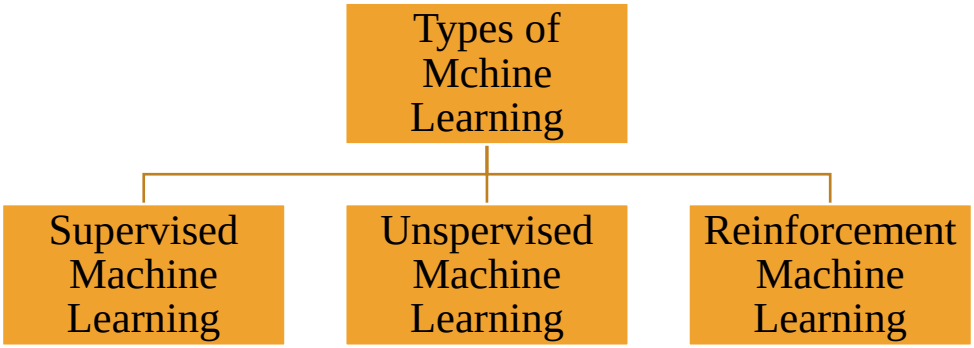



**Figure 1: Applications of the Machine Learning in Various Field.**

Studying the knowledge gained from prior experience and finding meaningful patterns from vast, complicated and unstructured data sets, the MAL uses optimization ranges, probabilities, and statistical techniques. Only a few applications for such algorithms have been developed: automated classification of texts, port scanning, electronic filters, credit card fraud detection, consumer payment detection, optimization in production processes and illness modelling[3].

Most of this application was developed related to an unattended machine learning algorithm[4][5]. There are three forms of master learning in the supervised version, based on the learning dataset with known tags, and the result of the unlabeled instances predicted below and displayed in the Figure 2.

- Unsupervised Machine Learning
- Supervised Machine Learning
- Reinforcement Machine Learning



Journal of The Gujarat Research Society

ISSN: 0374-8588
Volume 22 Issue 1, January 2020

**Figure 2: Types of ML: Supervised ML, Unsupervised ML, and Reinforcement ML.**

- *Unsupervised Machine Learning :*

Unattended learning takes place where the dataset has little or no interpretation. The labelling of the data is not defined. No other approach except via iterations is suitable to partition a data collection. In order to construct a structure by reviewing the relationship between the inputs, an input is used in supervised learning. Consider the categorization of animals, for example. Uncontrolled learning according to this research is not advised for prediction.

- *Reinforcement Machine Learning:*

Reinforcement learning's training of machine learnings model to makes sequences of decision. The agents learn to attain goals in uncertain, potentially complexes environment. In the reinforcement learning, artificial intelligences face game like situations. The computer employs error and trial to come with solutions to problems.

- *Supervised Machine learning:*

Approximations to functions can be characterized as controlled learning, and example to the generations of functions. If learning using the right training sets is done, well composed functions may be anticipated. The rise of supervised education based on data is steady. This is an apprenticeship which can lead to biased learning. For example, if x is self-multiplying inputs and output, the functions supplied by the monitored learning are chi square. Two kinds of controlled learning algorithms exist[6].

- Classification.
- Regression.

1. *Regression Algorithm:*

Regression algorithm is the algorithm which is use to find the future value by training the previous data by using the machine learning which is done by the programming in the computer by using different languages of coding such as python and R.

*1.1.Types of Regression:*

Figure 3 shows the seven types of regression are briefly explained are given below:

- Lasso Regression
- Linear Regression
- Logistic Regression
- Ridge Regression
- Elastic Net Regression
- Polynomial Regression
- Stepwise Regression

Journal of The Gujarat Research Society

ISSN: 0374-8588
Volume 22 Issue 1, January 2020

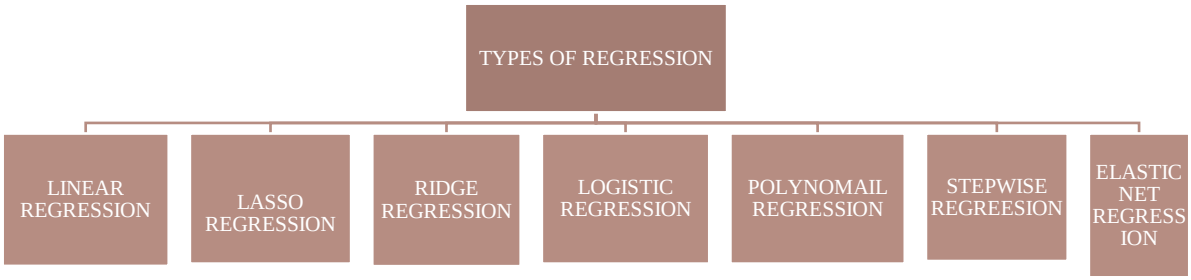



**Figure 3: Different Types of Regression Algorithms in Supervised Machine Learning**

- *Linear Regression:*

The most used modelling approach is linear regression. In this method, the first (dependent) variable is constant, and the second (dependent) variable is either discrete or continuous, resulting in a straight line that is the essence of that regression. It produces straight lines that best suit the computation of 1st (dependent) (y) and 1st (independent) (x) relationships of 3rd variables (which shows regression line). The regression line in the graph, as well as other data points around the regression line, are illustrated in Figure 4.

It is given by an equation:

$$y = c + m * x + e,$$

Where,

- c is intercept
- m is slope of the line and ,,
- And e is the error.

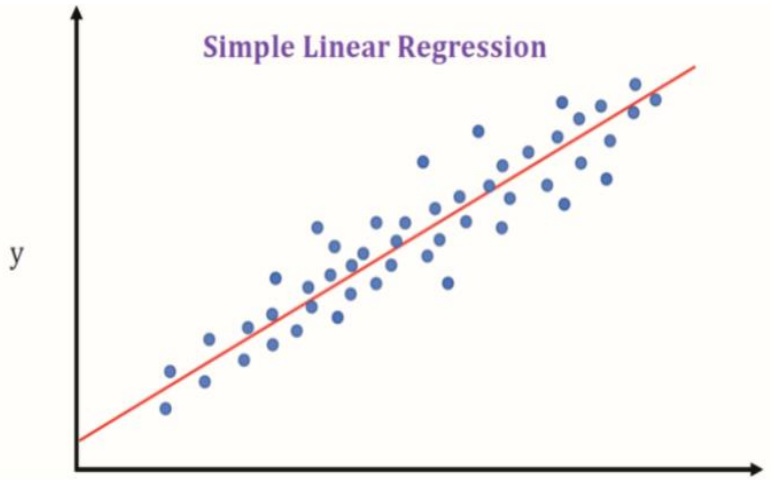



**Figure 4: Regression Line which is Plot in the Graph and Various Data Point Which is Close to the Prediction.**

Journal of The Gujarat Research Society

ISSN: 0374-8588
Volume 22 Issue 1, January 2020

- *Lasso Regression:*

Regression in Lasso is a kind of linear regression based on shrinking. Data values, like a mean, are compressed in shrinkage to the center spots. The lasso method promotes simple, sparse models. This regression approach was good for models with a high degree of multicollinearity or when specific models such as the elimination of variable parameters/selection must be automated. The least absolute operator for shrinkage and selection is a brief for the least actual operator of retraction and choice.

- *L1 Regularization:*

Regularization L1 is utilized for the regression of Lasso and the penalty must be equal to absolute values of coefficient magnitudes. This method of regularization might result in a sparse pattern of the few coefficients, with some coefficients 0 and the pattern eliminated. Larger penalty produces a factor of about 0, which is excellent to make a model simpler. L2, although in other respects, does not remove sparse or coefficient models. L2, however, does not. The Lasso is therefore far easier to grasp than the Ridge[7].

- *Ridge Regression:*

Ridge regression indeed model optimization technique which can used for analyses the data with the multicollinearity. L2 regularization achieved using that approach. Where they have a problem with the multicollinearity, least square are the unbiased, as well as the variances high, predicted values are so distant from the actual values.

The ridge regression cost function:

Minimum (|y – x (theta)|^2 + λ |theta|^2)

Where λ is penalty.

The ridge function's alpha parameter denotes the value given here.

Which can control the penalty term by adjusting the values of alpha. The greater the alpha value, the greater the penalty, and thus the magnitude of the coefficients' Decreased. It shrinks the parameters.

- *Logistic Regression:*

In the early 20$^{th}$ century, the biological sciences used logistic regression. It went on to be used in a variety of social science applications. Where the dependent variable (target) is categorical, logistic regression is used.

For instance,

- To predicts whether the emails are spam (1) or (0)
- To predicts whether the tumor malignant (1) ort (0)

Consider where you need to know whether email is a spam or not. In order to fix this problem, we must define a borderline by which data may be classified if linear regressions are used. If the value of the class is malignant, the predicted continuous score is 0.4 and the edge value is 0.5, the test prediction would be classified as not malignant, with serious implications.

Journal of The Gujarat Research Society

ISSN: 0374-8588
Volume 22 Issue 1, January 2020

It's clear from that example the linear regression is inappropriate for the classification problem. Meanwhile linear regression is unbounded, logistic regressions enter picture. This value is restricted to range of 1 to 0.

- *Polynomial Regression:*

Polynomial Regression is type of the regression investigation in that connection between the dependent and independent variables is modelled using a polynomial of n degree. There are lots of advantages of using polynomial regression some are given below. Figure 5 shows the polynomial curve which is plot in the graph and various data point which is close to the prediction**:**

- The best approximation of the connection between independent and dependent variable is polynomial.
- It may accommodate wide variability of the functions.
- Polynomial is the type of the curve which may match a large variability of curves.

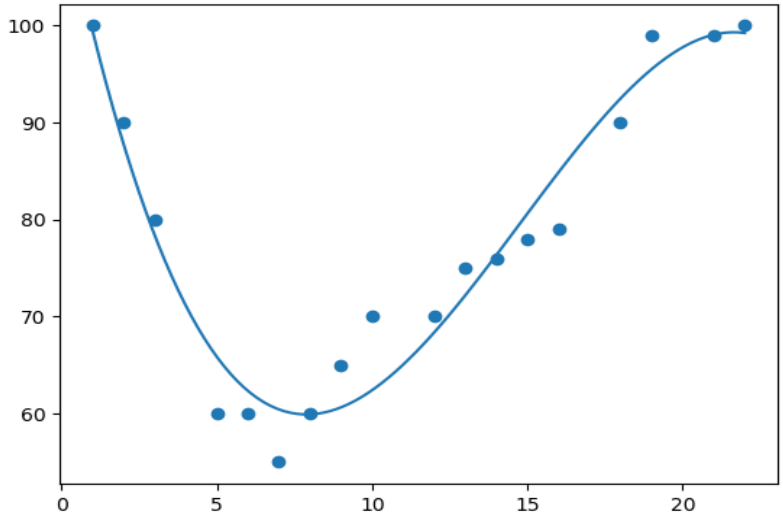



**Figure 5: Polynomial Curve which is Plot in the Graph and Various Data Point Which is Close To the Prediction:**

- *Stepwise Regression:*

The process of making models by deleting or adding a preview variable through a series of tests or an F-test step wise regression. By calculating the test statistics, the variable that has been omitted or added. Although the approach benefits, it requires skill on researchers' parts and can only be utilized by people who are knowledgeable in statistical testing. In the spirit, the step-by-step regression model with salt grains should be considered in contrast with the most regression model. The different benefits of step-by-step regression include:

- The ability to handle a great number of possible predictor variable while fine tuning model for choose best predictors variable from a pool of choices.
- It is quicker than the other methods of automated model selection.
- Observing order in which variables removed or the inserted will reveal useful information about the predictor variables' consistency.

Journal of The Gujarat Research Society

ISSN: 0374-8588
Volume 22 Issue 1, January 2020

- *Elastic Net Regression:*

The basic regression algorithm assumes a linear connection between the target and the inputs variable, which is known as linear regression. Addition penalties tools function during training encourages modest models with smaller coefficients value, which is an extension of linear regression. Regularized and penalize linear regression are terms used to describe these extensions.

Elastic net is a form of regularised linear regression that combines 2 commons penalties, the L2 and L1 penalty functions. After learning above, the summaries are given below:

- Elastic Net is a linear regression extension that includes regularization penalties in loss functions during the training.
- How to assess Elastic Net model as well as use finals model to forecast novel data.
- How to automatically customize Elastic Net models for novel dataset using grid scan.

## LITERATURE REVIEW

Tri Doan et.al studies one of the most popular tasks in data removal is to find the best algorithm for extracting important informations from data. As data excavation tools are more available, it cannot be feasible to test huge numbers of algorithms for the optimal algorithm on a single dataset. You illustrate how the trees may predict algorithm output in their papers by using multiple-variable linear regressions. They take the previous learning machine into account in building Meta Knowledge for supervised learning. The goal is to build this meta-knowledge by integrating the summary understanding of the information with the previous output of such datasets. They have shown that modified data sets generated by decreasing large dimensional spaces to smaller dimensions still retain key information needed for forecasting efficiency of the algorithms and complement clear statistical outlines with order to deliver the best and costs of categorization. Their approach works well enough for the nominal as well as the numerical data acquired in reality[8].

Iclhan Uysal et al. studies in the Statistical literature, learning or numerical features predicting are referred to as regression and are the subject of research in machine and statistics. Their study analyses the major algorithms of regression and the methods that both groups create. Many issues in the actual world may be treated as a regression problem and hence a regression is essential for the different applications. Locally Weight Regression, rule base regression, Projections Pursuit Regression, instance based regression, Multivariate Adaptive Regression Spline, as well as recursive partition regression method that induces regression tree are all discussed in their study[9].

Charalampos Bratsas et al. studies Researchers examined the prediction effectiveness of three machine learning models that use the probe data acquired from the Thessaloniki transport system: Random Forest, Support Vector Retrence and Multilayer Perceptron's and Linear Regressions Retrence. The comparison includes several tests, divided into 3 scenario types. The 1st stage tests the algorithms on certain randomly selected dates and highways. For eight successive fifteen minute periods the algorithms are tested arbitrarily in second state; the algorithms are tested for a whole day on randomly selected streets in third state. The Multilayer

*ISSN: 0374-8588*
*Volume 22 Issue 1, January 2020*

Perceptron model is best performed in stable settings with minor fluctuations. It is superior to scenarios that are more variable and has almost zero errors, depending on the outcome[10].

## DISCUSSION

The study examines everything about the supervised learning machine algorithm, which component of the machine learning is the regression algorithm. This report examines first the basics of machine learning as interpretation or concept of machine learning which specifies that machine learning learns the computer algorithm that further automatically improves through experience, and afterwards the different type of machine learning, unsurprising ML and reinforcement. Machine learning. ML discusses the review following that the ML algorithm is used in various applications, along with email filters, medicines and computer visions where traditional algorithms emerge, to accomplish the necessary tasks which are impossible or difficult, where it is difficult or impossible to develop traditional algorithms to accomplish the tasks required. After this article, we will examine the kind of regression method in supervised machine learning which is used to determine the prediction value in the form of prior data.

## CONCLUSION

After researching and analyzing this article, the machine learning that part of artificial intelligence utilizes increases every day, because the prediction of the future value of past data or training of previous data is predictable. The study also concludes that different forms of regression algorithms are useful for identifying future values or forecasted results, using crucial information in future forecasts. Machine learning in the 21st century is the most promising vocations. They have several excellent chances for payment job. In addition, the potential range of machine learning is changing the area of automation drastically. The possibilities of machine learning in India also exist. Machine learning therefore has a good life or reach in THE ML and contributes to the growth of the digital world.

**REFERENCES**

[1] X. Qiu, L. Zhang, Y. Ren, P. Suganthan, and G. Amaratunga, "Ensemble deep learning for regression and time series forecasting," *IEEE SSCI 2014 - 2014 IEEE Symp. Ser. Comput. Intell. - CIEL 2014 2014 IEEE Symp. Comput. Intell. Ensemble Learn. Proc.*, 2014, doi: 10.1109/CIEL.2014.7015739.

[2] J. Alzubi, A. Nayyar, and A. Kumar, "Machine Learning from Theory to Algorithms: An Overview," *J. Phys. Conf. Ser.*, vol. 1142, no. 1, 2018, doi: 10.1088/1742-6596/1142/1/012012.

[3] S. K. J. Vatsal Babel, Brijesh KumarSingh, "Evaluation Methods for Machine Learning," *J. Anal. Comput.*, vol. XI, no. I, pp. 1–6, 2018.

[4] J. Verrelst *et al.*, "Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel-2 and -3," *Remote Sens. Environ.*, vol. 118, pp. 127–139, 2012, doi: 10.1016/j.rse.2011.11.002.

[5] M. S. Mahdavinejad, M. Rezvan, M. Barekatain, P. Adibi, P. Barnaghi, and A. P. Sheth, "Machine learning for internet of things data analysis: a survey," *Digit. Commun. Networks*, vol. 4, no. 3, pp. 161–175, 2018, doi: 10.1016/j.dcan.2017.10.002.

[6] J. W. Baker, "Measuring bias in structural response caused by ground motion scaling," *Pacific Conf. Earthq. Eng.*, no. 056, pp. 1–6, 2007, doi: 10.1002/eqe.

[7] S. Safdar, S. Zafar, N. Zafar, and N. F. Khan, "Machine learning based decision support systems (DSS) for heart disease diagnosis: a review," *Artif. Intell. Rev.*, 2018, doi: 10.1007/s10462-017-9552-8.

[8] T. Doan and J. Kalita, "Selecting Machine Learning Algorithms Using Regression Models," *Proc. - 15th IEEE Int. Conf. Data Min. Work. ICDMW 2015*, pp. 1498–1505, 2016, doi: 10.1109/ICDMW.2015.43.

*ISSN: 0374-8588*
*Volume 22 Issue 1, January 2020*

[9] I. Uysal and H. A. Güvenir, "An overview of regression techniques for knowledge discovery," *Knowl. Eng. Rev.*, vol. 14, no. 4, pp. 319–340, 1999, doi: 10.1017/S026988899900404X.

[10] A. Dasgupta, Y. V. Sun, I. R. König, J. E. Bailey-Wilson, and J. D. Malley, "Brief review of regression-based and machine learning methods in genetic epidemiology: The Genetic Analysis Workshop 17 experience," *Genet. Epidemiol.*, vol. 35, no. SUPPL. 1, pp. 5–11, 2011, doi: 10.1002/gepi.20642.