

# Linear Regression in Machine Learning

B. Ganapathy Subramaniam<sup>1\*</sup>, T. Rama Prabha<sup>2</sup>

<sup>1</sup>Research Scholar, Software Development, CEI India Private Limited, Chennai, Tamil Nadu, India.

Email: giri.ganesh@gmail.com

<sup>2</sup>Department of Computer Science, VICAS, Tiruchengode, Tamil Nadu, India.

Email: mam.ramaradha1976@gmail.com

\*Corresponding Author

**Abstract:** Today, there are a lot of researches about prediction which turns to be nothing but a concept of regression is all about as a part of Machine Learning. Machine Learning works heavily on natural language processing, computational biology, computer vision, robotics and other areas. Techniques of machine learning are very important to understand the data to apply the algorithms to answer questions based on the data. Regression describes the relationships between various variables. When looking at the words that describe the relationships in people, being independent or dependent, being simple or complex, being positive or negative and also being strong or weak. This turns out that all the adjectives being used are directly related to regression analysis. Regression analysis enhances the machine learning by reducing the errors of prediction and makes the estimated values closer to the actual values. This paper briefly explains about the Machine learning and its most important technique called Linear regression with Gradient Descent.

**Keywords:** Gradient descent, Linear regression, Machine learning.

## I. INTRODUCTION

One area of computing which is continuously improving the way to use smartphones and web is Machine Learning. Sometimes Machine Learning and Artificial Intelligence get used to interchange between big brand companies to announce their latest innovation. However, Machine Learning (ML) and Artificial Intelligence (AI) are quite two different areas of computing. And of-course they are connected each other.

The goal of AI is to create the machine that mimic the human mind and to do that it needs learning capabilities. It is more than just about learning. Learning must be represented by knowledge representation, logical reasoning and abstract thinking. Machine Learning on the other hand, solely focus on writing software that can learn from the past experiences.

Modern world is all about humans and computers. Main difference on learning aspect between them are: Humans learn

from their past experience and computers need to be told what to do which needs to be programmed to perform any action by following specific instructions. Computers past experience is called "Data".

Quite an astounding fact is that Machine Learning is closer to Data mining as statistics than it is to Artificial Intelligence. One of the Standard definitions of Machine learning by Tom Mitchell from Carnegie Mellon University is, *A computer program is said to learn from experience "E" with respect to some class of tasks "T" and performance measure "P", if its performance at tasks in "T", as measured by "P", improves with experience "E" [3].*

In simple words, if a computer program can improve a performance of a certain task based on the past experience, then it proves that it has learnt.

## II. ROLE OF MACHINE LEARNING ALGORITHM

Machine Learning algorithm takes some sort of input in the form of data which are typically numerical values and produces the output which makes sense of data by predicting the result. Machine Learning algorithms produce two kinds of output, namely, Classification and Regression.

Classic example is a house price prediction. Series of data consisting of number of bedrooms, number of floors and area of the house will be given as an input to Machine Learning algorithm to predict its price as an output. For an example, data consists of 3 bedrooms, 2 floors and 1000 sq.ft is given as an input, output would be like Rs.10,00,000. Parameters 3, 2 and 1000 of the house which are called as data points are fed into an algorithm to get the price output. This is known as Regression which refers to predicting continuous numeric output. Whereas Classification refers to taking set of input and classifying into discrete set of labels. Based on the specific parameters in the house example, house can be classified as beach house, fancy house, guest house etc., Actually, output would be classified by specifying 90% beach house and 10% guest house. This is a classification attempting to assign a discrete label to a data point.

Few of the frequently used Machine Learning algorithms are k-Nearest Neighbour which works for both classification and

regression, Support Vector Machines and Artificial Neural Networks. Role of Machine Learning algorithm is to accept the unknown data point as an input and get a classification label as an output. But how could a Machine Learning algorithm performs the operation from unknown data. To make Machine Learning algorithm work and predict the result from unknown data, it must be trained. There are several different strategies that how machine can learn in order to perform operations. 1. Supervised learning, 2. Unsupervised learning and 3. Reinforcement learning.

Reinforcement learning is a kind of learning where the agent observes the environment and chooses an action. When performed action works well then there is a positive point, when it does not work then there is a negative point. Over the course by receiving more positive points, algorithm will perform better.

Unsupervised learning [4] is a learning generally applied to completely unknown data or the facts. Typically, unsupervised learning will be applied to the problem like clustering. By having the huge data and figuring out how these data be arranged into a group can be done by unsupervised learning.

Supervised learning [4] always goes with training data, test data and unknown or real time data. Huge set of input data with all the required parameters along with the actual output data often called as targets are used as training data. Training will be done over and over again to make algorithm work better. Once trained, test data will be injected in to verify the results to make sure the sufficient training. Test data is similar to training data but this should not be used to train the algorithm but to be used to verify the output. Once algorithm performs well with the test data, then it can be exposed to the real world of unknown data.

### III. LINEAR REGRESSION

Under a supervised learning, considering the above said classic example, let's have a classification of data by small house and big house priced 3 lakhs and 10 lakhs respectively. Task for an algorithm is to predict the price of the house based on its size. Let's try to estimate the medium sized house. To make this prediction, let's put the available data in the grid. Let's have the sizes of the house represented in square feet in the x axis and y axis represents the price of the house in Rupees. Let's take a look into the previously collected data in the below figure in the form of dots which represents the size of the house and its corresponding prices. Considering the training data, what would be best price for the medium sized house represented in the graph.

To figure out the best result, look at the formation of the data points which lays a path to plot a line. Best fit line according to the price of the house will guide to predict the price of the medium sized house as shown in the Fig. 2. Linear regression explores the relationship between a quantitative response variable and one or more explanatory variables. In simple words, Plotting the best fit line in regard to the independent variable and dependent variable is called as Simple Linear regression.

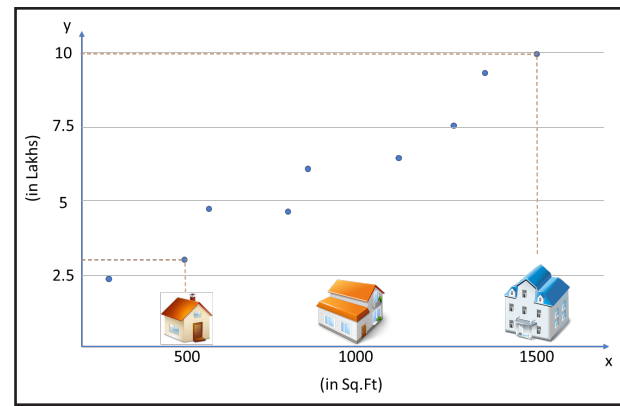


Fig. 1

Independent variable is in the x axis and dependent variable is in the y axis. This forms a relationship between two variables which allows to draw a straight line. When independent variable and dependent variable increases then it is a positive relationship. Let's say data has study time on independent variable and marks acquired on dependent variable. When study time increases, marks will also increase. There is a positive relationship. On the other hand, when independent variable increases and dependent variable decreases then it is a negative relationship. Let's put Facebook time instead of study time on independent variable. When Facebook time increases, marks will decrease which shows negative relationship.

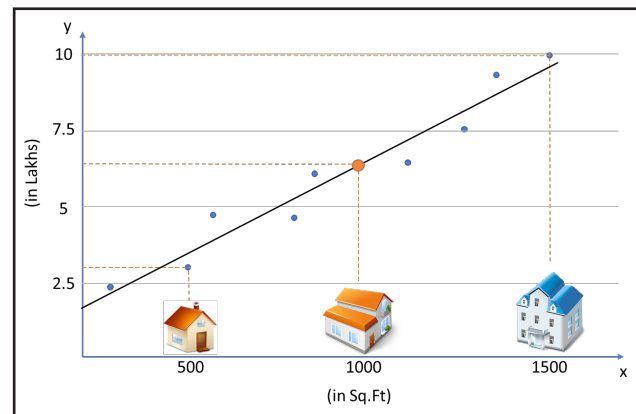


Fig. 2

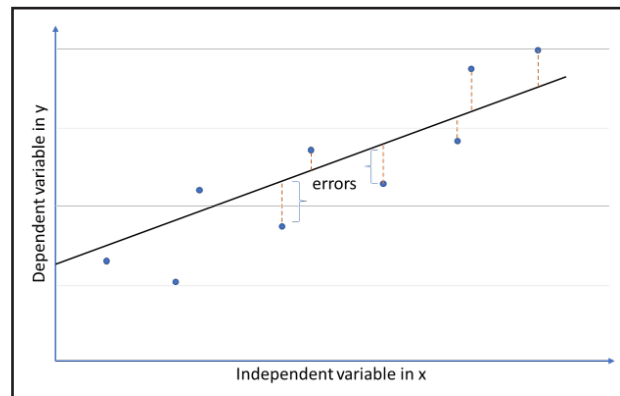


Fig. 3

Method used to draw the line is called Least Squares method and the line drawn is called regression line. And in the end, algorithm tries to minimize the errors between the estimated value and the actual value. Difference between the estimated value and actual value is called as Errors. Goal is to minimize these errors into as small as possible. A standard approach to minimize the errors is to define an error function or cost function which measures the best fit of the given line.

As in Machine Learning, machine has to be taught to figure out the best line to fit the data points. To achieve this, algorithm should use a standard straight-line equation  $y = mx + b$  where  $m$  is the line's slope,  $b$  is the y axis intercept,  $x$  and  $y$  are the coordinates of the data points. Error or Cost function will take  $(m, b)$  pair as a parameter to evaluate the error value and returns the sum of error values which represents the best fit of the line on the given data. To compute error on the given line, algorithm iterates through the given  $(x, y)$  data points and sums the square distances between each data point's  $y$  value and drawn line's  $y$  value. Purpose of squaring the distance is to make the value stay positive and make the error function differentiable.  $N$  represents the number of data points in the data set. Error function or cost function looks like:

$$\text{Error}_{(m,b)} = \frac{1}{N} \sum_{i=1}^N (y_i - (mx_i + b))^2$$

Best fitting line has lower error values.

#### IV. GRADIENT DESCENT

If the dataset is too homogenous, the training model will be too fit for the given data points. It will be over fit, meaning the training model will not be able to generalize the error. So at any given different data points, model will not be able to make an accurate prediction. Keeping the training data diverse is one of the important point to have better training model. One of the most important machine learning technique is Gradient Descent.

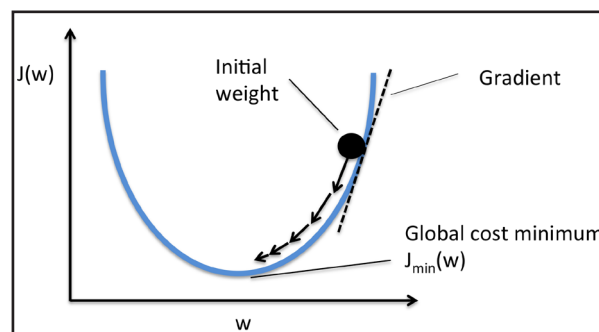
Gradient Descent is the foundation of how to train the intelligent systems based on the simple idea of immediately guessing the best solution to a given objective. Initial solution will be guessed and iteratively step in a direction closer to a better solution. Algorithm just repeats the process until it arrives at the solution that is good enough. Since there is no way to know the best solution from the start, it is easy to start with an educated guess and check method is extremely useful.

Purpose of executing gradient descent is to start from some location on the surface and move down to the plain with the lowest error. To evaluate the Gradient Descent on the Error function, gradient must be computed. Gradient will act as follow-through line which points to the target closer. Partial differentiation must be applied to Error function to compute the gradient. Since there are two different parameters  $(m, b)$ , partial derivative must be computed for each parameter. Derivatives looks like:

$$\frac{\partial}{\partial m} = \frac{2}{N} \sum_{i=1}^N -x_i(y_i - (mx_i + b))$$

$$\frac{\partial}{\partial b} = \frac{2}{N} \sum_{i=1}^N -(y_i - (mx_i + b))$$

Algorithm initializes the search to start at random pair values of  $(m, b)$  values and gradient descent algorithm works on Error( $J$ ) function to find out the best fit line. This will be repeated and on every iteration the value of  $(m, b)$  will be updated to a line that yields slightly lower error than the previous iteration. Learning Rate is an additional variable which is used to control the speed of the flow or a step during each iteration. It makes sure the algorithm not to skip over the minimum. On the occasion of smaller steps, algorithm requires many iterations to figure out the minimum to find the best fit line.



#### V. CONCLUSION

Idea of machine learning is to enhance the raw data A to sensible information B. Availability of lot of data today, there are lot of techniques to make the machine automatically learn from A to B. From the giant set of emails, there is an easy way to classify the spam. From the 50,000 hours of videos in YouTube, there is easy way to transcript the audio to text. Machine learning does a pretty good job by just mapping between A and B. Despite of all techniques in machine learning, it's still extremely limited on relative to the human intelligence yet machine learning drives the incredible amount of economic value. Understanding the concepts and implementations of Linear regression along with Gradient descent leads the way for better machine learning by means of speed of training, predictive accuracy, interpretability and memory usage.

#### REFERENCES

- [1] N. Altman, and M. Krzywinski, "Points of significance: Simple linear regression," *Nat. Methods*, vol. 12, no. 11, pp. 999-1000, 2015.
- [2] M. Andrychowicz, *et al.*, "Learning to learn by gradient descent by gradient descent," *NIPS*, pp. 1-16, 2016.
- [3] T. Mitchell, "Human and machine learning," *Machine Learning*, 2006.
- [4] A. Ng, "Supervised learning," *Mach. Learn.*, pp. 1-30, 2012.