A More Accurate Approach for Prediction using Gradient Descent

Shweta Agrawal
Professor I.A.C. S.A.G.E.
University Indore
M.P. India

Ravishek Kumar Singh
M Tech 4th Semester
Department of C.S.E. S.I.R.T
M.P. India

ABSTRACT
Accuracy is one of the most important concerned when we are dealing the problem related to machine learning. Artificial intelligence of also one of the most popular and emerging field which used optimization to improve accuracy. Gradient descent is a new and most power techniques which integrate machine learning and AI to achieve optimization. There are so many techniques are available for optimization and Gradient descent is one of then it is a kind of an iterative algorithm and most used for discovering the minimum cost function. This technique help to take efficiently and successfully decisions, by the use of derivatives. Derivative is useful when we want to calculate slope of the graph from a particular point. The value of the slope is defined by representation a tangent line to the graph at the point. In this technique we have to calculate this tangent line, with help of this we are able to calculate and decide the direction to reach the minima. It is the best techniques for optimization in machine learning. It is based on first-order optimization. This technique used a objective function, in which we have to update parameter in the reverse direction for every iteration. In the paper apply this optimization techniques to find better linear regression line and try to fit as best regression line for given data set. Our objective is to reduce SSE error and improve cost function value.

Keywords
Keywords — Machine Learning, Linear Regression, Optimization, Accuracy, Efficiently

1. INTRODUCTION
First we need to know concepts from linear regression. A linear regression model effort to describe the relationship between an output variables and one or more predictor variable using a straight line. This straight line is denoted using the following formula[7,8]:

$$y = mx + c$$

Where
y: dependent variable
x: independent variable
m: Slope of the line
c: y intercept

The initial step is to define a linear regression equation which we need to determine, if there is a relationship between the two variables. We can used concepts of correlation coefficient and by apply scatter plot. Correlation coefficient shows that we are able to predict outcomes and a scatter plot shows that data appears to form a straight line. We used simple linear regression model to find a predictive function. Let us consider an example. In this example we have taken data for spent in marketing and sales[12,13].

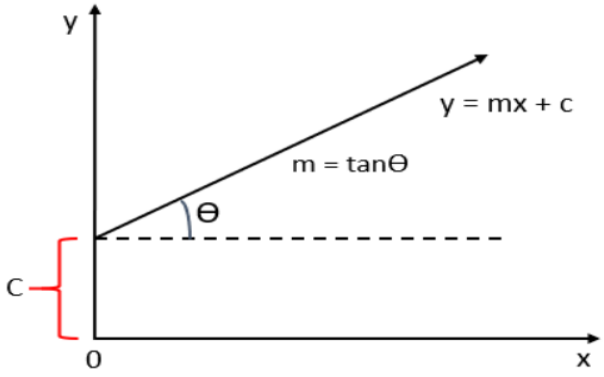

Figure 1 Linear regression model

Where sales are dependent variable and spent on marketing is independent variable.

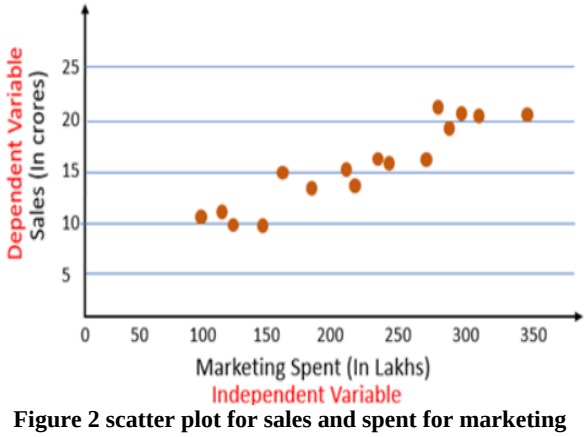

Figure 2 scatter plot for sales and spent for marketing

From the scatter plot it is clearly shown that there is a linear relationship between dependent variable and independent variable.

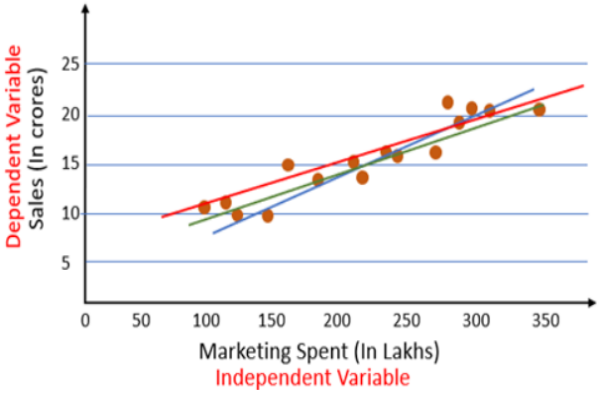

Figure 3 multiple lines are passes through these points

Now we need to find a straight line between dependent variable and independent variable that explain the relationship between

them. But we can see that there are multiple lines that can pass through these points.
So how do we know which of these lines is the best fit line? That's the problem that we will solve in this article. For this, we will first look at the cost function.

2. GRADIENT DESCENT APPROACH

Gradient descent is very common and best recognized techniques which are used for optimization. This approach is based on iterative procedure to complete optimization. This technique used a cost effective function the value of this function is need to be minimize. This process helps to make decision proficiently and offers best solution to the problem. This used derivatives; derivative is a significant term that originates from calculus. Derivative is applied to compute the slope for a particular point. The tangent describes the slop by depiction the graph for a particular point. We have to compute tangent by calculating value of tangent, we are able to find the desired direction and easy reach for the minima. This approach takes only the first derivative and performs the updates for the parameters. We can define the term as gradient for a partial derivative applied for inputs [11,12,14].

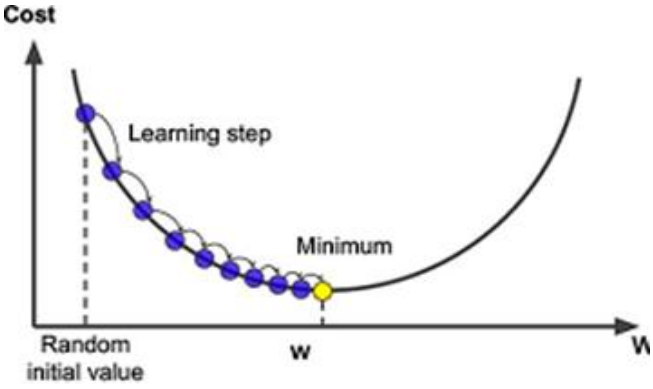

Figure 4 process of gradient descent approach

The Gradient Descent approach defines the values of m and c, such that line equivalent to those values is best fitting line with minimum error. First we have to compute the loss function. This function gives us the error in our predicted value of m and c. Our objective is to minimize this error to obtain most accurate value of m and c. To compute the loss, we need to find out MSE (Mean Squared Error) .we need to do following things

1. Find out the difference between actual and predicted value of dependent variable.
2. Find out the square of difference.
3. ·Find out the mean of squares for every value of independent variable.

Now, we have to minimize "m" and "c". To minimize these parameters, Gradient Descent Algorithm is used.

3. LITERATURE SURVEY

In 2018 Simon S. Du et al proposed "Gradient Descent Learns One-hidden-layer CNN: Don't be Afraid of Spurious Local Minima" They took a problem of learning a one-hidden layer of in which both the convolutional weights w and the output weights are parameters to be learned. They proved using Gaussian input Z, spurious local minimizer. They also proved the first polynomial convergence guarantee can randomly initialized for learning a one-hidden-layer convolutional neural network. They t revealed an interesting occurrence that randomly modified local search algorithm can converge to a global minimum or a spurious local minimum. They gave a quantitative characterization of gradient descent dynamics to explain the two-phase convergence phenomenon. Experimental results also verified by theoretical findings [1].

In 2018 Shadi Diab et al proposed "Optimizing Stochastic Gradient Descent in Text Classification Based on Fine-Tuning Hyper-Parameters Approach". Their objective was to improve performance of (SGD) for text classification. They proposed SGD learning with Grid-Search method to fine-tuning hyper-parameters for enhancing the performance of SGD classification. They explored different settings for representation, transformation and weighting features of terrorist attacks incidents obtained from the Global Terrorism Database. They concluded that using a grid-search performance of the classifiers and execution time are improved. They applied SVM, Logistic Regression and Perceptron classifiers to classify the terrorist attacks. They explored different experiments to compare the performance of the classifiers without SGD learning, with SGD learning and with SGD learning including hyper-parameter discovery [2].

In 2018 Lala Septem Riza et al proposed " Development of *R* Package and Experimental Analysis on Prediction of The co2 Compressibility Factor Using Gradient Descent". They focused on building a package written in R that includes eleven procedures based on gradient descent. They showed the accuracy and computational cost are reasonable for the average of root mean square root and simulation time. They developed the *R* package, namely "gradDescent". They implemented a linear model based on gradient descent with regression. They showed that the procedures included in the package provide reasonable predicted values. The package "*gradDescent*" is alternative software for dealing with various regression tasks in the realistic world problems[3].

In 2018 Nan Cui, proposed "Applying Gradient Descent in Convolutional Neural Networks". They discussed the CNN and the related BP and GD procedures. They included the basic structure and function of CNN, details of each layer, the principles and features of BP and GD. They introduced the Convolutional Neural Network process, which is very common in machine learning and image recognition. They combined the Back Propagation mechanism and the Gradient Descent method, at the same time, the practical application and are discussed. With the constant development of technology, the neural network is bound to become an important tool in people's lives. For the neural network, the major challenges in the future may not only exist in the more improved system, but also in the quality and quantity of the training set[4].

In 2019 Jiawei Zhang "Gradient Descent based Optimization Algorithms for Deep Learning Models Training". They provided an introduction to the gradient descent based optimization algorithms for learning deep neural network models. Deep learning models involving multiple nonlinear projection layers are very challenging to train. Nowadays, most of the deep learning model training still relies on the back propagation algorithm actually. In back propagation, the model variables will be updated iteratively until convergence with gradient descent based optimization algorithms. Besides the conventional vanilla gradient descent algorithm, many gradient descent variants have also been proposed in recent years to improve the learning performance, including Momentum, Adagrad, Adam, Gadam, etc., which will all be introduced in this paper respectively[5].

In 2019 Dr. Yvonne W. Karanja et al proposed "Review On Gradient Descent Algorithms In Machine Learning" Deep learning (DL) is assuming an inexorably significant They used

some streamlining strategies to improve the exactness of the research and diminish preparing time**.** They showed that developing impressive methods, the subject is logically drawing. They attempted and introduce a thorough audit on earlier work led in deep learning; there are remaining parts to improve the learning cycle. They showed that instance of the current spotlight is loaning fruitful thoughts from different territories of machine learning. They explained setting of dimensionality decrease and much work that should have been done. They brought a more profound jump into the unique training algorithms and architectures[6].

In 2019 Christian L. Thunberg et al proposed "Stochastic Gradient Descent in Machine Learning". They recognized program and implemented with the parameters of the stochastic gradient descent. They analyzed how effect on the computation speed and accuracy. The implemented digit recognizing system for accuracy and tested time per iteration stayed constant. They showed at low learning rates yielded a slower rate of convergence while larger ones yielded faster but more unstable convergence. They improved the convergence for computational power. They used new parameters of stochastic gradient descent and analyzed in regards to their effect on the neural network optimization, in particular the parameters and for learning rate[7].

In 2020 Abdelkrim El Mouatasim et al proposed "Fast Gradient Descent Algorithm for Image Classification with Neural Networks". They proposed modified version of gradient descent technique and the learning rate is update in each epoch. They learned learning rate by Armijo rule. They called this approach as fast gradient descent (FGD). Proposed algorithm is applied on MNIST dataset. By the experiment, they showed that FGD approach is much faster than gradient descent algorithms. In the proposed work the gradient descent algorithm is accelerated with Nestrov step and learning rate via Armijo condition. They proved that convergence rate of FGD algorithm is quadratic convergence, and the comprising of numerical results[9].

In 2020 M. Raja et al proposed "Stochastic Gradient Descent with Logistic Regression Technique for Medical Data Classification". They presented a new stochastic gradient descent approach with logistic regression. They presented a model to execute data pre-processing as the initial stage, perform conversion and class labeling. They developed a new model for disease diagnosis. By the outcome of the model they denoted that the SGDLR-MDC model has proficient performance with the maximum accuracy on the applied hepatitis, ILP, and thyroid dataset respectively[10].

In 2020 Yanli Liu et al proposed "An Improved Analysis of Stochastic Gradient Descent with Momentum". They applied SGD with momentum (SGDM) in many machine learning tasks. The momentum weights tuned in a stage wise manner. They showed that SGDM converges fast as compared to SGD for smooth objectives under both strongly convex and non-convex settings. They proved that multistage strategy is beneficial for SGDM compared to using fixed parameters. They verified theoretical claims by numerical experiments Multistage SGDM. They established its convergence and show the advantage of stage wise training. They addressed all still open problems. For example, (a) Is it possible to show that SGDM converges faster than SGD for special objectives such as quadratic ones[11].

In 2021 Peshawa Jamal Muhammad et al proposed "Gradient Descent Algorithm: Case Study". They clarified the mathematical background of the gradient descent approach with a simple case study. The approach is one of the most important optimization strategies used by many machine learning models. This algorithm is widely used in deep learning, neural networks, linear regression, and logistic regression. They simply presented that how the algorithm is worked with a case study; the model function determined that represents the collected data. Gradient descent is one of the efficient algorithms that can be used in finding variant optimum values of the modeling equations[15].

In 2021 Babacar Gayeet al proposed "Sentiment classification for employees reviews using regression vector stochastic gradient descent classifier (RVSGDC)". They proposed a new approach based on lexicon-based and machine learning techniques. They extracted the sentiments of employees from dataset as positive and negative. They proposed a hybrid and voting model denoted as Regression Vector-Stochastic for sentiment classification. This model combined logistic regression, support vector machines, and stochastic gradient descent. They also used other machine learning models in the performance comparison. They extracted term frequency-inverse document frequency and global vectors are used to train learning models. They also evaluated the performance of all models in terms of accuracy, precision, recall, and F1 score[13].

4. STEPS USED IN GRADIENT DESCENT

The below pseudo-code is a modified version from the source:
1. First initialize the m and c with random values
For example m = 1.1326 and c =0.0224, line equation is y = mx+b, then y = 1*x+2. This can give a line that is not best fitted for the historical data, as shown below.

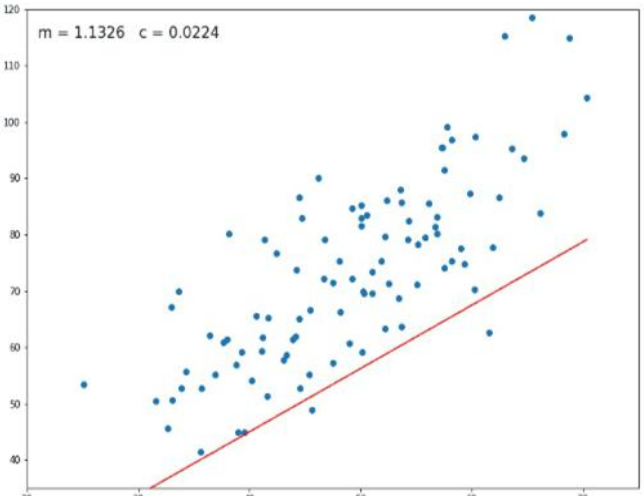

Figure 5 Initial parameters for regression line

The line is not best fitted, and the idea here is to find the optimal values of the line's coefficients that can best fit the historical data. Once we find the best line that fit the data, we can use this regression model for prediction.

2. Calculate the gradient.

Use the Sum of Squared Errors (SSE) to find cost function to minimize the prediction error. The gradient of SSE is a partial derivative of SSE

SSE equation is as follows:

$$SSE = \frac{1}{2}(Y - \bar{Y})^2$$

$$\bar{Y} = mx + b$$

Derivative of SSE

$$\frac{\partial}{\partial m} SSE = -(Y = \bar{Y})X$$

Derivative of SSE

$$\frac{\partial}{\partial m} SSE = -(Y = \bar{Y})$$

The gradient is made up with partial derivatives i.e. both partial derivatives of SSE as shown below:

$$\nabla SSE = \begin{bmatrix} \frac{\partial}{\partial m} SSE \\ \frac{\partial}{\partial m} SSE \end{bmatrix} = \begin{bmatrix} -(Y = \bar{Y})X \\ -(Y = \bar{Y}) \end{bmatrix}$$

In the first step, we initialize m and c with random values. For the subsequent iterative process, m and c values are updated using step 3.

3. Update coefficients value of optimal m and c.

Update the coefficients m and c using the gradient calculated from the above equations.

$$m = m - r * \frac{\partial}{\partial m} SSE$$

$$b = b - r * \frac{\partial}{\partial m} SSE$$

Here r is the learning rate.

4. Use new m and b for prediction

With new m and c, compute in the above step, to draw the line that fit the data. Again calculate the SSE of each data point in X to find out the total SSE.

5. Repeat steps 2, 3 and 4

Repeat the steps 2, 3 and 4 until optimal values for the coefficients m and b are found that reduces the SSE to a minimum value.

We found the best line that fit the data for our regression model using the gradient descent algorithm.

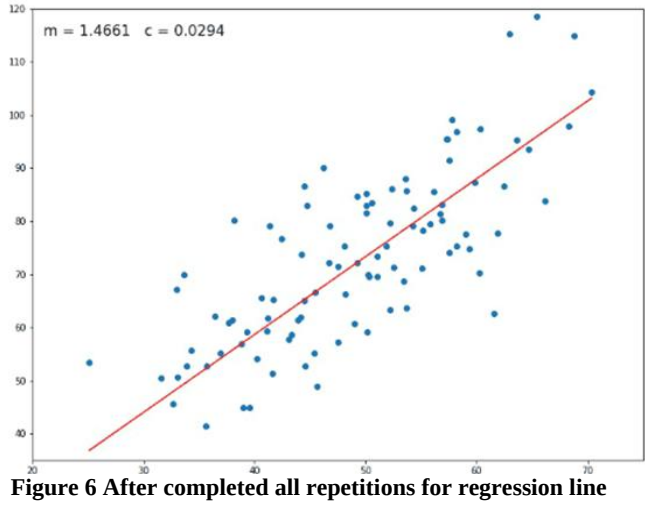

Figure 6 After completed all repetitions for regression line

5. CONCLUSION

There are so many techniques are available for optimization and Gradient descent is one of then it is a kind of an iterative algorithm and most used for discovering the minimum cost function. This technique help to take efficiently and successfully decisions, by the use of derivatives. Derivative is useful when we want to calculate slope of the graph from a particular point. The value of the slope is defined by representation a tangent line to the graph at the point. In this technique we have to calculate this tangent line, with help of this we are able to calculate and decide the direction to reach the minima. It is the best techniques for optimization in machine learning. It is based on first-order optimization. This technique used an objective function, in which we have to update parameter in the reverse direction for every iteration. In the paper apply this optimization techniques to find better linear regression line and try to fit as best regression line for given data set. Our objective is to reduce SSE error and improve cost function value. We used real life data set for house size and house price. first we fit correct regression line and the we are able to predict price for a for a given size of a house. Proposed approaches predict better and correct value as compared to other techniques.

6. REFERENCES

[1] Simon S. Du , Jason D. Lee and Yuandong Tian Barnab´as P´oczos 1 Aarti Singh Gradient Descent Learns One-hidden-layer CNN: Don't be Afraid of Spurious Local Minima Proceedings of the 35 th International Conference on Machine Learning, Stockholm, Sweden, PMLR 80, 2018. Copyright 2018.

[2] Shadi Diab, Al-Quds Optimizing Stochastic Gradient Descent in Text Classification Based on Fine-Tuning Hyper-Parameters Approach International Journal of Computer Science and Information Security (IJCSIS), Vol. 16, No. 12, December 2018.

[3] Lala Septem Riza, Dendi Handian, Rani Megasari, Ade Gafar Abdullah, Asep Bayu Dani Nandiyanto, Shah Nazir4 Development Of R Package And Experimental Analysis On Prediction Of The Co2 Compressibility Factor Using Gradient Descent Journal of Engineering Science and Technology Vol. 13, No. 8 (2018) 2342 - 2351 © School of Engineering, Taylor's University.

[4] Nan Cui1, Applying Gradient Descent in Convolutional Neural Networks 1Electrical and Computer Engineering, University of Massachusetts Lowell, Lowell,01854, United States IOP Conf. Series: Journal of Physics: Conf. Series 1004 (2018) 012027 doi :10.1088/1742-6596/1004/1/012027.

[5] Jiawei Zhang Gradient Descent based Optimization Algorithms for Deep Learning Models Training Founder and Director Information Fusion and Mining Laboratory (First Version: February 2019; Revision: March 2019.) arXiv:1903.03614v1 [cs.LG] 11 Mar 2019.

[6] Dr. Yvonne W. Karanja Review On Gradient Descent Algorithms In Machine Learning Electronic copy available at: https://ssrn.com/abstract=3810947 .

[7] Christian L. Thunberg Niklas Mannerskog Stochastic Gradient Descent in Machine Learning Degree Project in Technology, First Cycle, 15 Credits Stockholm, Sweden 2019 ,

[8] Dmitrii Marin Meng Tang Ismail Ben Ayed Beyond Gradient Descent for Regularized Segmentation Losses 1https://github.com/dmitrii-marin/adm-seg 2In this paper, "shallow" refers to methods unrelated to deep learning 2019.

[9] Abdelkrim El Mouatasim Fast Gradient Descent Algorithm for Image Classi_cation with Neural Networks Abdelkrim El Mouatasim Ibn Zohr University, Faculty of Poly

disciplinary Ouarzazate (FPO), B.P. 284, Ouarzazate 45800, Morocco. E-mail: a.elmouatasim@uiz.ac.ma 2020.

[10] M. Rajaa, M. Y. Mohamed Parvees Stochastic Gradient Descent with Logistic Regression Technique for Medical Data Classification International Conference on IoT based Control Networks and Intelligent Systems (ICICNIS 2020) Department of Computer and Information science, Annamalai University, mraja.cdm@gmail.com.

[11] Yanli Liu_ Yuan Gaoy Wotao Yin An Improved Analysis of Stochastic Gradient Descent with Momentum 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

[12] Karen E. Walker, Walker Consulting LLC Chandler Arizona Gradient Descent Using SAS™ for Gradient Boosting PharmaSUG 2020 - Paper XX-25.

[13] Babacar Gaye, Dezheng Zhang and Aziguli Wulamu Sentiment classification for employees reviews using regression vectorstochasticgradient descent classifier (RVSGDC) Babacar Gaye, Dezheng Zhang and Aziguli Wulamu 2021. Sentiment classification for employees reviews using regression vector-stochastic gradient descent classifier (RV-SGDC). PeerJ Comput. Sci. 7:e712.

[14] Chuanlei Zhang , Minda Yao , Wei Chen , Shanwen Zhang , Dufeng Chen,and Yuliang Wu6 "Gradient Descent Optimization in Deep Learning Model Training Based on Multistage and Method Combination Strategy" Accepted 2 July 2021; Published 23 July 2021 Academic Editor: Chi-Hua Chen Copyright © 2021 Chuanlei Zhang.

[15] Peshawa Jamal Muhammad Ali and Haval Abdulkarim Ahmed Gradient Descent Algorithm: Case Study Machine Learning Technical Reports is a periodical technical report published by the Machine Learning Laboratory Koya 44023, Erbil, F.R. of Iraq