Volume 9, Issue 08, August 2024 ISSN (online): 2456-0006

International Journal of Science Technology Management and Research

Available online at: www.ijstmr.com

Gradient Algorithm an Improved Approach for Prediction Using Regression Model

Triveni Jangde
M Tech CSE 4th sem
LNCT (Bhopal) Indore Campus
Indore M.P. India

Dikshika Maliwad
Assistant Professor CSE department
LNCT (Bhopal) Indore Campus
Indore M.P. India

Abstract:* Predictive analytics i.e. forecasting future opportunities and risks is the most prominent application of Gradient descent algorithm. Demand analysis, for instance, predicts the number of items which a consumer will probably purchase. However, demand is not the only dependent variable when it comes to business. RA can go far beyond forecasting impact on direct revenue. we can forecast the number of shoppers who will pass in front of a particular billboard and use that data to estimate the maximum to bid for an advertisement. Insurance companies heavily rely on regression analysis to estimate the credit standing of policyholders and a possible number of claims in a given time period. Data Science understanding is key for Gradient descent algorithmis not only great for lending empirical support to management decisions but also for identifying errors in judgment. For example, a retail store manager may believe that extending shopping hours will greatly increase sales. Gradient descent algorithm, however, may indicate that the increase in revenue might not be sufficient to support the rise in operating expenses due to longer working hours (such as additional employee labor charges).In this paper we uisng Gradient descent Predictive analytics

***Keywords:* Predictive, Analytics, Gradient, Descent, Forecast, Regression**

I. INTRODUCTION

Gradient descent is an optimization algorithm used to find the values of parameters (coefficients) of a function (f) that minimizes a cost function (cost). Gradient descent is best used when the parameters cannot be calculated analytically (e.g. using linear algebra) and must be searched for by an optimization algorithm. Some of the reasons behind gradient descent algorithm that's why we motivated towards this technique. Optimization is a big part of machine learning. Gradient descent is a simple optimization procedure that you can use with many machine learning algorithms. Batch gradient descent refers to calculating the derivative from all training data before calculating an update. Stochastic gradient descent refers to calculating the derivative from each training data instance and calculating the update immediately. Optimization may be defined as the process by which an optimum is achieved. It is all about designing an optimal output for problems with the use of resources available. Optimization in machine learning is slightly different. In most of the cases, we are aware of the data, the shape and size, which also helps us know the areas we need to improve. But in machine learning we do not know how the new data may look like, this is where optimization acts perfectly. Optimization techniques are performed on the training data and then the validation data set is used to check its performance. There are a lot of advanced applications of optimization which are widely used in airway routing, market basket analysis, face recognition and so on. Machine learning algorithms such as linear regression, KNN, neural networks completely depend on optimization techniques. Here, we are going to look into one such popular optimization technique called Gradient Descent[1,2,3].

II . APPLICATION OF GRADIENT DESCENT ALGORITHM

A Predictive Analytics:

Predictive analytics i.e. forecasting future opportunities and risks is the most prominent application of Gradient descent algorithm. Demand analysis, for instance, predicts the number of items which a consumer will probably purchase. However, demand is not the only dependent variable when it comes to business. RA can go far beyond forecasting impact on direct revenue. For example, we can forecast the number of shoppers who will pass in front of a particular billboard and use that data to estimate the maximum to bid for an advertisement. Insurance companies heavily rely on regression analysis to estimate the credit standing of policyholders and a possible number of claims in a given time period. Data Science understanding is key for predictive analytics[4,5,6].

B Operation Efficiency

Gradient descent algorithmcan also be used to optimize business processes. A factory manager, for example, can create a statistical model to understand the impact of oven temperature on the shelf life of the cookies baked in those ovens. In a call center, we can analyze the relationship between wait times of callers and number of complaints. Data-driven decision making eliminates guesswork, hypothesis and corporate politics from decision making. This improves the business performance by highlighting the areas that have the maximum impact on the operational efficiency and revenues[9,10].

C Supporting Decisions

Businesses today are overloaded with data on finances, operations and customer purchases. Increasingly, executives are now leaning on data analytics to make informed business decisions that have statistical significance, thus eliminating the intuition and gut feel. Gradient descent algorithm can bring a scientific angle to the management of any businesses. By reducing the tremendous amount of raw data into actionable information, Gradient descent algorithm leads the way to smarter and more accurate decisions. This does not mean that Gradient descent algorithm is an end to manager's creative thinking. This technique acts as a perfect tool to test a hypothesis before diving into execution[7,8].

D Correcting Errors

Gradient descent algorithm is not only great for lending empirical support to management decisions but also for identifying errors in judgment. For example, a retail store manager may believe that extending shopping hours will greatly increase sales. Gradient descent algorithm, however, may indicate that the increase in revenue might not be sufficient to support the rise in operating expenses due to longer working hours (such as additional employee labor charges). Hence, this analysis can provide quantitative support for decisions and prevent mistakes due to manager's intuitions[11,12].

III. LITERATURE SURVEY

In 2017 Stephan Mandt and Matthew D. Hoffman proposed "Stochastic Gradient Descent as Approximate Bayesian Inference". Stochastic Gradient Descent with a constant learning rate (constant SGD) simulates a Markov chain with a stationary distribution. With this perspective, they derive several new results. (1) Constant SGD can be used as an approximate Bayesian posterior inference algorithm. Specifically, how to adjust the tuning parameters of constant SGD to best match the stationary distribution to a posterior, minimizing the Kullback Leibler divergence between these two distributions(2) They demonstrate that constant SGD gives rise to a new variation EM algorithm that optimizes hyperparameters in complex probabilistic models. (3) They also showed how to tune SGD with momentum for approximate sampling. (4) They analyze stochastic-gradient MCMC algorithms. For Stochastic-Gradient Langev Dynamics and Stochastic-Gradient Fisher Scoring, we quantify the approximation errors due to finite learning rates. (5) They used stochastic process perspective to give a short proof of averaging is optimal. They proposed scalable approximate MCMC algorithm, the Averaged Stochastic Gradient Sampler[13].

In 2017 Sebastian Ruder proposed "An overview of gradient descent optimization algorithms". Gradient descent optimization algorithms while increasingly popular are of ten used as black-box optimizers, as practical explanations of their strengths and weaknesses are hard to come by. They provide the reader within tuition with regard to the behavior of different algorithms that will allow her to put them to use. In the course of this overview, different variants of gradient descent, summarize challenges, introduce the most common optimization algorithms, review architectures in a parallel and distributed setting, and investigate additional strategies for optimizing gradient descent. They initially looked at the three variants of gradient descent, among which mini batch gradient descent is the most popular. They investigated algorithms that are most commonly used for optimizing SGD: Momentum, Nesterov accelerated gradient, Adagrad, Adadelta, RMSprop, Adam, AdaMax, Nadam, as well as different algorithms to optimize asynchronous SGD.Finally, they considered other strategies to improve SGD such as shuffling and curriculum learning,batch normalization, and early stopping[14].

In 2017 Shuang Song Kamalika Chaudhuriand Anand D. Sarwate proposed"Stochastic gradient descent with differentially private updates". Differential privacy is a recent framework for computation on sensitive data, which has shown considerable promise in the regime of large datasets. Stochastic gradient methods are a popular approach for learning in the data-rich regime because they are computationally tractable and scalable. They derive differentially private versions of stochastic gradient descent, and test them empirically. Results showed that standard SGD experiences high variability due to differential privacy, but a moderate increase in the batch size can improve performance significantly. Mini batched stochastic gradient descent (SGD). When details plentiful, privacy is "affordable," and SGD strategies are more computationally efficient. They showed that in many cases the performance of differentially private SGD was close to that of non-private SGD, especially with larger batch sizes. Instochastic optimization, both the variability of the algorithm and the impact of privacy-preserving noise can be ameliorated by processing groups of points together. By experiments showed that privacy affects both the optimal batch size b and learning rate. Some interesting future directions suggested by the work include[15].

In 2018 Loucas Pillaud-Vivien and Alessandro Rudi proposed "Statistical Optimality of Stochastic Gradient Descent on Hard Learning Problems through Multiple Passes". They considered stochastic gradient descent (SGD) for least-squares regression with potentially several passes over the data. While several passes have been widely reported to perform practically better in terms of predictive performance on unseen data, the existing theoretical analysis of SGD suggests that a single pass is statistically optimal. While this is true for low-dimensional easy problems, they showed that for hard problems, multiple passes lead to statistically optimal predictions while single pass does not; they also showed that in these hard models, the optimal number of passes over the data increases with sample size. In order to define the notion of hardness and show that our predictive performances are optimal, they consider potentially infinite-dimensional models and notions typically associated to kernel methods, namely, the decay of eigenvalues of the covariance matrix of the features and the complexity of the optimal predictor as measured through the covariance matrix. They illustrated results on synthetic experiments with non-linear kernel methods and on a classical benchmark with a linear model[16].

In 2018 Leon Bottou FrankandE. Curtis Jorge proposed "Optimization Methods for Large-Scale Machine Learning". They provide a review and commentary on the past, present, and future of numerical optimization algorithms in the context of machine learning applications. Through case studies on text classification and the training of deep neural networks. They discussed how optimization problems arise in machine learning and what makes them challenging. A major theme of our study is that large-scale machine learning represents a distinctive setting in which the stochastic gradient (SG) method has traditionally played a central role while conventional gradient-based nonlinear optimization techniques typically falter. They present a comprehensive theory of a straightforward, yet versatile Algorithm, discuss its practical behavior, and highlight opportunities for designing algorithms with improved performance. This leads to a discussion about the next generation of optimization methods for large-scale machine learning, including an investigation of twomain streams of research on techniques that diminish noise in the stochastic directions andmethods that make use of second-order derivative approximations[17].

IV. PROPOED APPROACH

Each of the gradient descent algorithms has their strengths and weaknesses. Steps used a proposed approach

- For rapid prototyping, adaptive techniques like Adam/Adagrad getting quicker results with much less efforts apply require much hyper-parameter tuning.
- We should apply gradient descent or momentum. Gradient descent is slow to get the desired results, but these results are mostly better than adaptive techniques.
- If data is small and can be fit in a single iteration, we can use 2nd order techniques like l-BFGS. This is because 2nd order techniques are extremely fast and accurate, but are only feasible when data is small enough
- There also an emerging method (which I haven't tried but looks promising) to use learned features to predict learning rates of gradient descent.

V. RESULT ANALUYSIS

when the value of m=1.1326 and value of c=0.0224

This screenshot display regression line when the value of m=1.1326 and value of c is 0.0224. From the screen shot we can see that the line is for way from all the give data set and it is not best fitted line for the given data set.

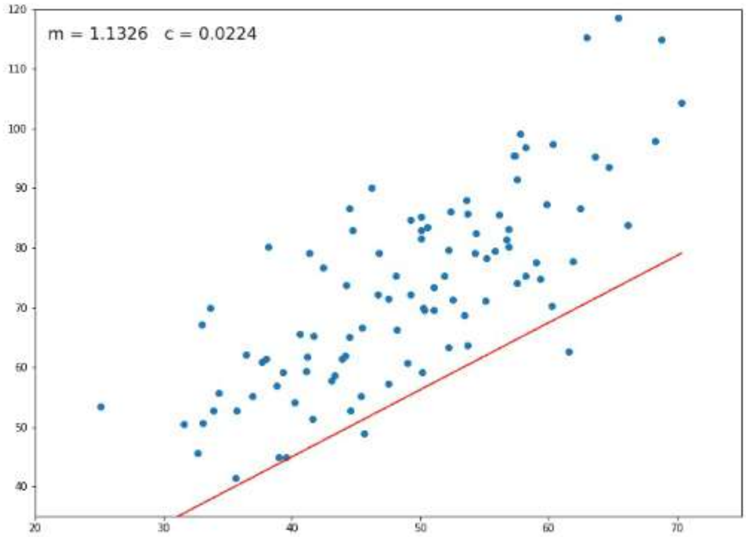

Fig 1 Regression line when the value of m=1.1326 and value of c=0.0224

when the value of m=1.2547 and value of c=0.0249

This screenshot display regression line when the value of m=1.2547 and value of c is 0.0249. From the screen shot we can see that the line is for way from all the give data set but it is improved regression line as compared to previous one .

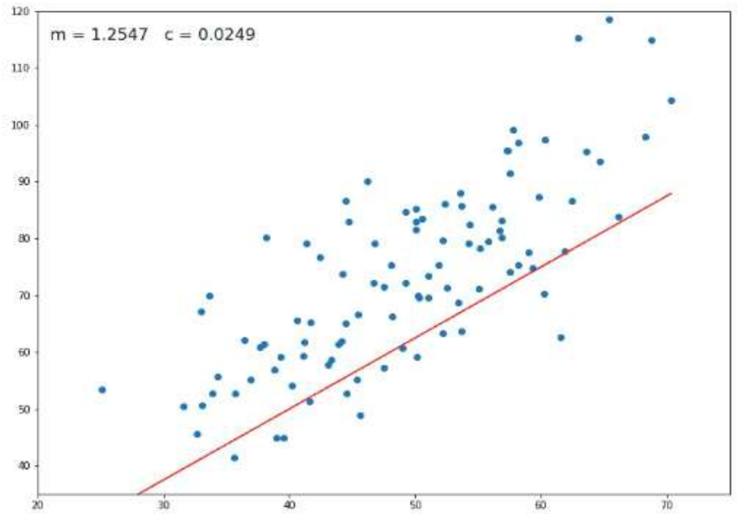

Fig2 Regression line when the value of m=1.2547 and value of c=0.0249

when the value of m=1.4591 and value of c=0.0292

This screenshot display regression line when the value of m=1.4591 and value of c is 0.0292. From the screen shot we can see that the line is more and more near from all the give data set but it is good fitted regression line as compared to previous one .

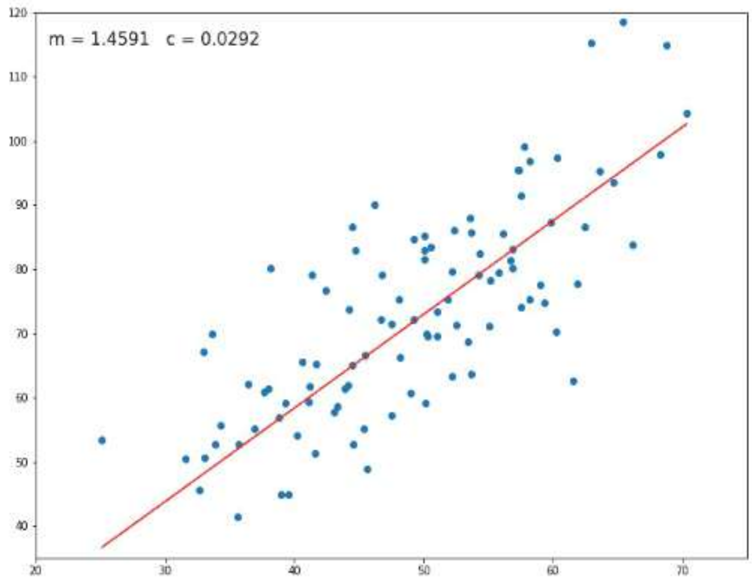

Fig 3 Regression line when the value of m=1.4591 and value of c is 0.0292

CONCLUSION

Gradient Descent Algorithm helps us to make these decisions efficiently and effectively with the use of derivatives.In the proposed work we used Gradient descent for Optimization we apply random value for m and c. We continue apply iteration and we found that after some iteration we found best fitted regression line for the given data set. For practical implementation we used house size and house price data set. By the experimental analysis we found that the proposed approach give better regression line for the proposed data set and gets reduce error up to remarks.

REFERENCE

1. Jonathan Schmidt, Mário R. G. Marques , Silvana Botti Recent advances and applications of machine learning in solidstatematerials science 26 February 2019 Accepted: 17 July 2019.
2. Simon Shaolei Du Gradient Descent for Non-convex Problems in Modern Machine Learning APRIL 2019 CMU-ML-19-102 Machine Learning Department School of Computer Science Carnegie Mellon University Pittsburgh, PA 15213.
3. Yura Malitsky Konstantin Mishchenko Adaptive Gradient Descent without Descent Proceedings of the 37 th International Conference on Machine Learning, Vienna, Austria, PMLR 119, 2020.
4. Nam D. Vo , Minsung Hong and Jason J. Jung Implicit Stochastic Gradient Descent Method forCross-Domain Recommendation System Sensors 2020, 20, 2510; doi:10.3390/s20092510 www.mdpi.com/journal/sensors Western Norway Research Institute, Box 163, NO-6851 Sogndal, Norway

5. Yoshimasa TsuruokaJunichi TsujiiSophia Ananiadou "StochasticGradient Descent Training for L1-regularized Log-linear Modelswith Cumulative Penalty"Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP, pages 477–485, Suntec, Singapore, 2-7 August 2009. c 2009 ACL and AFNLP.
6. Leon Bottou Large-Scale Machine Learningwith Stochastic Gradient Descent Y. Lechevallier, G. Saporta (eds.), Proceedings of COMPSTAT'2010,DOI 10.1007/978-3-7908-2604-3 16, c Springer Verlag Berlin Heidelberg 2010.
7. Feng Niu, Benjamin Recht, Christopher Hogwild A Lock-Free Approach to Parallelizing StochasticGradient Descent Computer Sciences Department, University of Wisconsin-Madison1210 W Dayton St, Madison, WI 53706 June 2011.
8. Matthew D. Zeiler, ADADELTA: AN ADAPTIVE LEARNING RATE METHOD arXiv:1212.5701v1 Dec 2012.
9. Ilya Sutskever James Martens "On the importance of initialization and momentum in deep learning" Proceedings of the 30 th International Conference on Machine Learning, Atlanta, Georgia, USA, 2013. JMLR: W&CP volume 28. Copyright 2013 by the author(s).
10. Yiming Ying and Massimiliano Pontil Online gradient descent learning algorithm Department of Computer Science, University College London Gower Street, London, 2014 WC1E 6BT, England, UK fying, m.pontilg@cs.ucl.ac.uk.
11. Diederik P. Kingma Jimmy Lei Ba ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION arXiv:1412.6980v9 [cs.LG] 30 Jan 2017.
12. Marcin Andrychowicz, Misha Denil, Sergio Gómez Colmenarejo1 Learning to learn by gradient descent by gradient descent. Western Norway Research Institute, Box 163, NO-6851 Sogndal, Norway
13. Stephan Mandt and Matthew D. Hoffman Stochastic Gradient Descent as Approximate Bayesian Inference Journal of Machine Learning Research 18 (2017) 1-35 Submitted 4/17; Revised 10/17; Published 12/17.
14. Sebastian Ruder An overview of gradient descent optimizationalgorithmsarXiv:1609.04747v2 . Insight Centre for Data Analytics, NUI Galway Aylien Ltd., Dublinruder.sebastian@gmail.com
15. Shuang Song Kamalika Chaudhuri Stochastic gradient descent with differentiallyprivate updates Dept. of Computer Science and EngineeringUniversity of California, San DiegoLa Jolla, CA USA shs037@eng.ucsd.edu.
16. Loucas Pillaud-Vivien Alessandro Rudi Statistical Optimality of Stochastic Gradient Descent on Hard Learning Problems through Multiple Passes INRIA - Ecole Normale Supérieure PSL Research University loucas.pillaud-vivien@inria.fr.
17. Leon BottouFrank E. CurtisJorge Nocedal Optimization Methods forLarge-Scale Machine Learning SIAM REVIEW 2018 Society for Industrial and Applied Mathematics Vol. 60, No. 2, pp. 223–311.