IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Compassion between Various optimization techniques used in Machine Leaning

1Rohit Garg, 2Sohit Aggarwal

1M. Tech student Department of Computer Science, 2HOD, Department of Computer Science

1Suresh Gyan Vihar University, India,

2Suresh Gyan Vihar University, India

ABSTRACT

Optimization in simplest terms optimization is choosing inputs that will result in the best possible outputs or making things the best that they can be. In the era from 2006 to 2009 peoples are looking to improve learning algorithms. At that time the best learning algorithm was gradient descent and the researchers are looking for a better learning algorithm than gradient descent. The researchers are looking for a better solution to the initialization problem in a neural network or some clear wait configuration. And that time they have sigmoid as a nonlinear activation function and they are searching for some better activation function. So they have started to invent new optimization techniques to improve all above parameters. Here in this paper we are comparing some latest Optimisation algorithms and made the conclusion that which algorithm is better in artificial intelligence e applications.

Keywords:- Artificial Intelligence, Optimization Algorithms, Adam, Nestrov, Gradient Descent

I. INTRODUCTION

One way to think about machine learning is as a way to convert raw data to a simplified cartoon-like representation a model and to use that representation to make predictions and estimates whenever we're engaged in the first part fitting a model to some data there is some optimization [1-2] going on it's built into the foundation of machine learning it comes to the foreground particularly when we're training deep neural networks.

Optimization can be applied to a huge variety of situations and problems for example choosing the optimal location for a warehouse to minimize shipment times to potential customers, designing a bridge that can carry the maximum load [3-4] possible for a different cost choosing the optimal build order for units, in a strategy game to amass the strongest army possible in a given time controlling the insulin output from an artificial pancreas to minimize the difference between actual and desired blood sugar levels throughout the day, designing an airplane wing to minimize as we can see optimization is a powerful tool in many applications this is just a small sampling of the many fields that make use of optimization techniques to improve the quality of their solutions if something can be modeled mathematically it can usually be optimized to summarize optimization improves results by helping to choose inputs that produce the best outputs most optimization problems require an optimization algorithm to solve and optimization is applicable to many disciplines.

II. VARIOUS OPTIMIZATION ALGORITHMS

Here we are comparing some optimizations and techniques on deep learning and we all know that optimization algorithm is the main approach used today for training a machine learning or deep learning model to minimize its error rate. There are two metrics to determine the efficiency of an optimizer the first one is the speed of convergence and the second one is the generalization popular algorithms such as ADAM or stochastic gradient [5] descent can capably cover one or our network but receivers can have it both ways. So we discuss about momentum and RMSprop [6] and ADAMs [7] that is the most popular techniques on deep learning in my opinion.

- Momentum

what is the momentum let's have the situation that instead of using the gradient to change the position of the weight let's use is to change the velocity if we have the cost function here in the mini-batch we have the same cost function similar like this as we can see it's much more frequency in time. It's classical theoretical and we have dissertation that imagine a ball on the error surface the location of the ball in the horizontal tile plane represents the weight vector the ball starts off by following the gradient but owns it has velocity it no longer does steepest descent it's momentum makes it keep on the previous

direction it damps oscillations in direction of high converters by combining ratings with opposite signs. Momentum is a matter that he has to accelerate stock - gasps the gradient descent and it's working like this it's accelerate its local minimum so momentum reduce updates for dimension who's great and change directions it's increased updates for dimensions whose greatest points in the same directions.

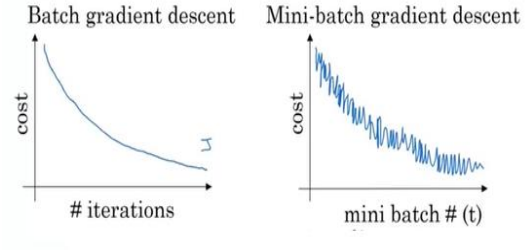

Figure.*1 Difference between batch gradient and mini batch gradient*

- RMSprop:

RMSprop dividing the learning rate for aweight by a running average of themagnitudes of recent ratings for thatweight it is a mini batch version of just using design of the gradient.

$$E[g^2]_t = 0.9E[g^2]_{t-1} + 0.1g_t^2$$
$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{E[g^2]_t + \epsilon}} g_t$$

RMSprop is a mini batch versionof LP prop.Prop is more popular and it can be usedin mini batch models which is also oneof the way how we can to optimize ourdeep learning model so we can go furtherto the intuition of RMS prop and weshould know that as RMS prop divides thelearning rate by an exponentiallydecaying average of square gradients andthe RMS probe is a rootmean square propagation and it's basedon exponentially weighted average. It's aformula is given blow and working isshown in Fig.2 if we have a low starting position shown

$$v(t) = Bv(t-1) + (1-B)\varphi(t)$$

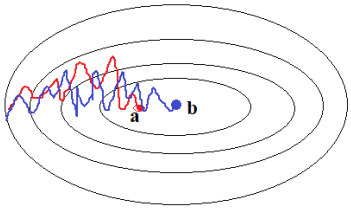

Figure. 2 Contour diagram for optimization

By red line local minimum will be at point 'a' the RMS probe work is like blue colour line to auscultation andlearning process.

- Adam

So after we know what is the RMSprop we'll go to the ADAM and what isthe ADAM is the combination ofstochastic great in the sense ofmomentum and as RMSprop and weremember that momentum accelerate our learning process and elseimprove is reflect to auscultation and this islike the combination ofacceleration and hostile reflection ofoscillations and ADAM is an adaptivemoment estimation which uses estimationof first and second moments of gradientsto adapt a learning rate for each weightof the new neural network and what isthe moment in here and moment of arandom variable is defined as theexpected value of that variable. Adam is an optimization algorithm thatcan use instead of classical stochasticgradient descent procedure to updatenetwork weights iterative based I average of the second moment of thegradients and this algorithm calculatesan exponential moving average of thegradient and square gradient and theparameter better one and better tocontrols that decay rates of this movingaverage and in conclusions.

III. RESULT AND DISCUSSION

In this paper work we have made a comparison in between various optimization algorithms under various numbers of hidden layers. Six different algorithms are used here with following network parameters.
Number of hidden layer = 100
Activation function is Soft Max.
Noise introduce = 0.01
We got the following result

Table 5.1 Algorithms vs mean accuracy and standard accuracy (100 hidden layers)

S. No	Algorithms	Mean Accuracy	Standard Accuracy
1	MOMENTUM	0.8618666666666668	0.018305069121845888
2	RMS PROP	0.8722666666666666	0.009826947078767051
3	ADAM	0.8781333333333333	0.001643843734125059 4

By the above table we can see that the accuracy of ADAM Algorithm is best among other two algorithm similarly Momentum having the lowest accuracy.

REFERENCES

1. *Sonia Chaliia "Opimization of Induced Drag with Varying Angle of Attack" Global Journal of Engineering Science and Researches. Vol. 6 Issue 5, 2019, pp 512-516*
2. *Gupta G. "Drowsiness Monitoring System Using Machine Learning", Global Journal of Engineering Science and Researches. Vol. 6 Issue 5, 2019, pp 507-511.*
3. *Budor Mohammed Abdelaiti "Stochastic Population Growth Model Using Matlab", Global Journal of Engineering Science and Researches. Vol. 7 Issue 1, 2020, pp 8-16.*
4. *S. Sivahsankar, Dr. R Jaganathan & D. Surendra. "Prediction And Prevention Of Crimes Using Geospatial Bigdata Analytics." Global Journal of Engineering Science and Researches. Vol. 6 Issue 2, 2019, pp 13-16.*
5. *Aditi Gupta "Direct face recognition: neural-network approaches." Global Journal of Engineering Science and Researches. Vol. 6 Issue 2, 2019, pp 358-361*
6. *Amandeep Kaur & Parveen Kakkar "Yawning classification framework for driver drowsiness detection.". Global Journal of Engineering Science and Researches. Vol. 6 Issue 5, 2019, pp 430-435.*
7. *Budoor Mohammed Abdelati "Stochastic Population Growth Model Using Matlab", Global Journal of Engineering Science and Researches. Vol. 7 Issue 1, 2020, pp 8-16.*