Effect of Learning Rate on Artificial Neural Network in Machine Learning

Igiri Chinwe Peace
Department of Computer Science
University of Port Harcourt
Nigeria

Anyama Oscar Uzoma
Department of Computer Science
University of Port Harcourt
Nigeria

Silas Abasiama Ita
Department of Computer Science
University of Port Harcourt
Nigeria

Abstract—**Machine learning has wide range of application in almost every life endeavor. Artificial neural network technique in particular has been used to implement prediction and forecasting of result in virtually all works of life including weather, sports, student performance etc. such parameters as momentum, training cycles and leaning rate play significant role in optimization of prediction or forecasting results. This research investigates the effect of learning rate in training a model using Artificial Neural Network technique. 15 iterative learning rates yielded an undulating graphical representation. The study further shows 80% prediction with at 0.1 learning rate and 90% prediction at 0.8 learning rate. This implies that applying the appropriate optimization strategy in machine learning could result the best possible result.**

Keywords—ANN; BNN; Machine Learning; Learning Rate; Prediction; Momentum

I. INTRODUCTION

The use of Artificial Neural Network (ANN) in modelling and predictive analysis has gained so much popularity over the years in various areas of science and human endeavour. The role, use and relevance of data set in these areas can never be over emphasized. Several Artificial Neural Network (ANN) and Biological Neural Network (BNN), which is a natural extension of the Artificial Neural Networks (ANN) have algorithms in different forms and shades to help address the problem domains that Artificial Neural Network (ANN) solves. Some of the algorithms are:

- Error correction Learning
- Gradient Descent
- Back propagation

The back propagation algorithm has been the most popular approach for neural networks training/classification due to its flexibility and robustness. This method has been used to solve used to solve various real life problems.

1) *What is NEURAL NETWORK?*

The term "Neural Network" can be referred to as both the natural and artificial alternatives, though classically this term is used to refer to artificial and external systems only. An Artificial Neural Network (ANN), usually called a neural network (NN), is a mathematical model or computational model that is inspired by the structure and functional aspects of biological neural networks [2]. Mathematically, neural nets are regarded as nonlinear objects with each layer representing non-linear combination/variations of non-linear functions from the prior layers. Each neuron in the network is a multiple-input, multiple-output (MIMO) scheme that receives pointers from the inputs, produces a subsequent signal, and communicates that signal to all possible outputs.

Basically, neurons in an Artificial Neural Network (ANN) are arranged into different discrete layers. The first and topmost layer is the one that interacts with the surroundings to receive various combinations of possible input is known as the input layer.

The last and final layer that interacts with the output to present the final processed data is known as the output layer.

While the layers that are between the input and the output layer that do not have any real communication with the environment are known as hidden layers. Hence increasing the complexity of an Artificial Neural Network (ANN), and its computational ability, requires the additions of a lot of more hidden layers and neurons per layer.
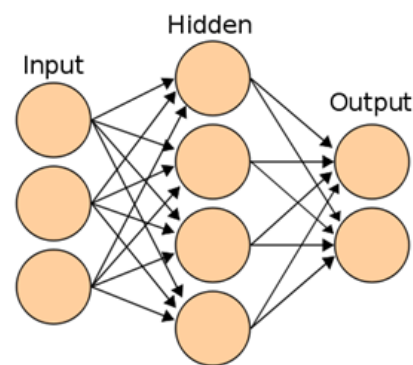

Fig. 1: Layers of Artificial Neural Network

2) *Neural Network Learning*

Learning is a very important module to every intelligent system. Looking at artificial neural network, learning typically happens during a precise training/classification phase. Once the neural network has been trained, it goes into a phase called the production phase where it produces results independently. In This phase, training can take on diverse forms, using a mixture of learning archetypes, learning guidelines, and learning algorithms. A network which has discrete learning and

production phases is referred to as a static network. Networks that are able to continue learning during production phase are known as dynamical systems. According to [3], neural network is a massively large parallel interconnection of several neurons that achieve different perception tasks in a small amount of time.

3) Learning Paradigm

A learning paradigm is supervised, unsupervised or a hybrid of the two that can reflect the method in which training data is presented to the neural network. A method that combines supervised and unsupervised training is known as a hybridized system. A learning rule is a model/concept that addresses the various types of methods to be used in the training of the system and also an area for what types of results are to be expected.

The ANN learning algorithm is a dedicated mathematical method that is used to update the neural weights during each of the training iteration. Each learning rule provides for a variety of possible learning algorithms to be used. A lot of ANN algorithms can only be used with a single learning rule. Learning rules and learning algorithms can typically be used with either supervised or unsupervised learning paradigms, however, and each will produce a different possible effect in the results. The process of building predictive models requires a well-defined training and validation protocol in order to ensure the most accurate and robust predictions. This kind of protocol is sometimes called supervised learning [1].

The learning rate is a problem that arises when very large training samples are provided, and the underlying system becomes incapable of valuable generalization. This can also occur when there are too many neurons in the network and the volume of computation exceeds the dimensionality of the input in the vector space.

During training, a lot of caution must be applied so as not to provide too many input examples and different numbers of training examples could produce very different results in the excellence and strength of the network.

II. RELATED WORKS

An Artificial Neural Network (ANN) which is a non-linear data modeling tool that can be used to find hidden patterns and relationships within data was developed by [4].

The following steps were employed: Data Collection, Data Extraction, and Data mining.

- Preliminary study using Total yardage differential, Rushing yardage differential, Time of possession differential (in seconds), Turnover differential, Home or away
- Training of data
- Parameter selection/classification
- Making predictions

The findings by the author show good solution to the prediction problem.

The major pitfalls include:

- High complexity,
- Lowly parameterized
- Poor training of datasets.

A learning rate of 0.01 was used.

The week 15 prediction rate was 75% using the season average and only 37.5% of the games using the three week average.

A hybridized approach using Linear Regression and Artificial Neural Network (ANN) was developed by [5]. Data mining steps were used in the investigation

The findings by the authors show good solution to the prediction problem.

The major pitfalls include:

- Low learning rate,
- Too many features
- Use of polynomial data.

A learning rate of 0.2 was used with a prediction accuracy of 90.3 percent obtained

A comparative approach using Logistic Regression and Artificial Neural Network (ANN) was developed by [6] using an Improved Prediction System for Football a Match Result. Proper data mining steps were used in the experiment by the author

The findings by the authors show good solution to the prediction problem.

The major pitfalls include:

- High learning rate,
- Too many in game features
- Use of binomial data.

A learning rate of 0.2 was used with a prediction accuracy of 85 percent obtained

III. THEORITICAL FRAMEWORK

A) Artificial Neural Network (ANN)

The system develops a model using a Back Propagation approach with emphasis on the learning rate for different prediction label. It further checks the resultant predictor effect on the different learning scheme employed by the network. A trained neural network can be thought of as an "expert" in the category of information it has been given to analyse. This expert can then be used to provide projections given new situations of interest and answer "what if" questions [8].

This technique learns a model by means of a feed-forward neural network trained by a back propagation algorithm (multi-layer perceptron). This operator cannot handle polynomial attributes. The results from will the linear regression will be converted using a nominal to numeric operator

1) Description

An artificial neural network (ANN), usually called neural network (NN), is a mathematical model or computational model that is inspired by the structure and functional aspects of biological neural networks. A neural network consists of an interconnected group of artificial neurons, and it processes information using a connectionist approach to computation (the central connectionist principle is that mental phenomena can be described by interconnected networks of simple and often uniform units) [5].

A feed-forward neural network is an artificial neural network where connections between the units do not form a directed cycle.

Back propagation algorithm is a supervised learning method which can be divided into two phases: propagation and weight update. The two phases are repeated until the performance of the network is good enough. In back propagation algorithms, the output values are compared with the correct answer to compute the value of some predefined error-function.

A multilayer perceptron (MLP) is a feed-forward artificial neural network model that maps sets of input data onto a set of appropriate output. An MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. Thus it can be used to perform a nonlinear prediction of a stationary time series, however, a time series is said to be stationary when its statistics do not change with time [7].

2) Algorithm (ANN)

Input

Attributes X1, X2…. Xn

Main process algorithm

initialize network weights (often small random values)

  do

    forEach training example

      prediction = neural-net-output(network, ex)

      actual = teacher-output(example)

      compute error (prediction - actual)

      compute   all weights from hidden layer to output layer

      compute all weights from input layer to hidden layer

      update network weights // input layer not modified by error estimate

    until all ex classified correctly or another terminating criterion satisfied

    return the network

3) Output

Y= Predicted Result used for rating

B) Analysis of theoretical framework

In designing the system, the following steps will be employed:

- Step 1: Problem definition

Here the understanding will be broken into the project requirements.

- Step 2: Data collection and pre-processing

- Step 3: Modeling

This phase is the core of the system and will be divided into two sub steps:

- Build

ANN will be developed using the features sets.

- Execute

LEARNING RATE	PREDICTION VECTOR
0.1	80.65
0.15	80.65
0.2	80.65
0.25	80.65
0.3	83.87
0.35	87.14
0.4	87.1
0.45	87.1
0.5	87.1
0.55	83.87
0.6	80.65
0.65	87.1
0.7	87.1
0.75	90.3
0.8	90.3

Table1: Learning rate versus prediction vector
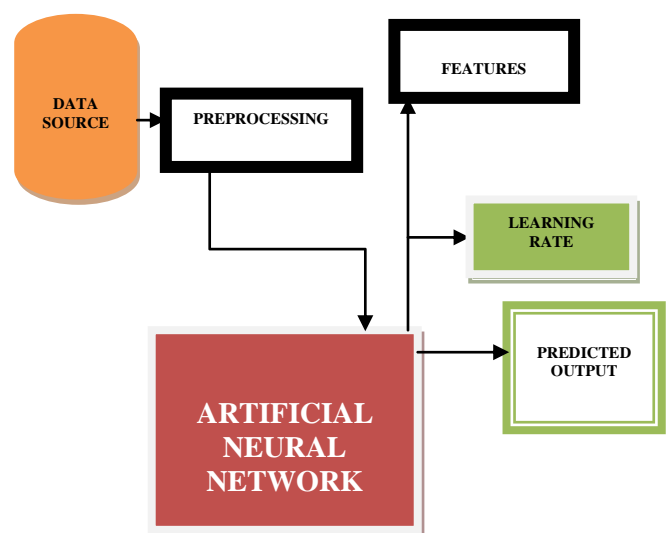
Fig. 2: Architectural Framework
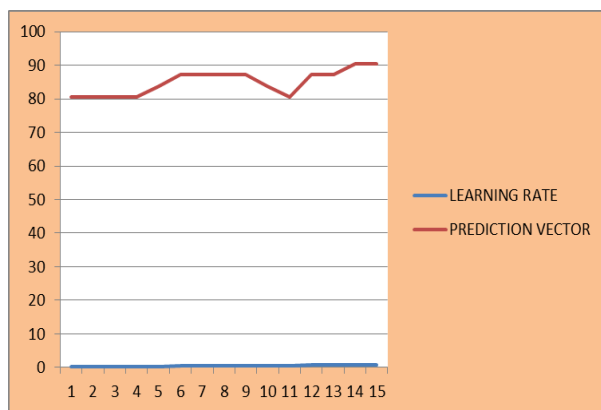
Fig. 3: Modelling Result

Fig 4: Graph showing Learning Rate

- OUTPUT

ImprovedNeuralNet

Hidden 1

========

Node 1 (Sigmoid)

Home Team = Denver Broncos: 1.027

Home Team = New York Jets: 0.777

Home Team = Buffalo Bills: 0.742

Home Team = St. Louis Rams: 0.583

Home Team = Carolina Panthers: 0.292

Home Team = Detroit Lions: 0.917

Home Team = Chicago Bears: 1.645

Home Team = Jacksonville Jaguars: 0.985

Home Team = Dallas Cowboys: 1.023

Home Team = Pittsburgh Steelers: -0.879

Home Team = Cleveland Browns: 1.849

Home Team = San Francisco 49ers: -0.031

Home Team = Indianapolis Colts: -0.109

Home Team = New Orleans Saints: -1.277

Home Team = Washington Redskins: 0.230

Home Team = San Diego Chargers: 0.081

Home Team = New England Patriots: -0.511

Home Team = Kansas City Chiefs: 0.515

Home Team = Philadelphia Eagles: 0.598

Home Team = Seattle Seahawks: -0.001

Home Team = Baltimore Ravens: 0.043

Home Team = Oakland Raiders: 0.110

Home Team = Atlanta Falcons: -0.434

Home Team = Arizona Cardinals: 0.314

Home Team = Houston Texans: 1.067

Home Team = Green Bay Packers: 2.223

Home Team = Tampa Bay Buccaneers: -0.814

Home Team = New York Giants: -0.065

Home Team = Cincinnati Bengals: -0.940

Home Team = Tennessee Titans: 1.591

Home Team = Minnesota Vikings: 0.301

Home Team = Miami Dolphins: 0.369

Away Team = Baltimore Ravens: 0.361

Away Team = Tampa Bay Buccaneers: 0.965

Away Team = New England Patriots: -0.736

Away Team = Arizona Cardinals: 0.931

Away Team = Seattle Seahawks: 0.279

Away Team = Minnesota Vikings: 0.063

Away Team = Cincinnati Bengals: -0.881

Away Team = Kansas City Chiefs: 0.558

Away Team = New York Giants: 1.997

Away Team = Tennessee Titans: 1.293

Away Team = Miami Dolphins: 0.531

Away Team = Green Bay Packers: 0.800

Away Team = Oakland Raiders: -0.468

Away Team = Atlanta Falcons: 0.395

Away Team = Philadelphia Eagles: 1.715

Away Team = Houston Texans: -0.710

IV.   RESULT DISCUSSION

From the results obtained, looking at the varying effect of the learning rate on the prediction vector, a graphical representation of the result is obtained using fifteen adjusted iterations. The graph shows an insignificant between 0.1 and

0.25 learning rate with respect to prediction of 80.65% when applied to a time series football result data. More so, there was a sharp increment from 83.87% to 87.14% when then learning rate was adjusted from 0.03 to 0.35. In a nutshell therefore, the fluctuation shows a low prediction of 80.65% with 0.1 learning rate, 80.65% at the learning rate of 0.6, and 90.3% at learning rate of 0.8 as shown in table 1 and fig. 5. It can be deduced from this result that a prediction system can be optimized with 0.8 factor learning rate at 500 training cycles and 0.2 momentum.

Thirty one seasonal vectors were used over the adjusted iterations, the several vectors that were incorrectly predicted by the Artificial Neural Network Model over the progression of the seasonal data represents the vectors are not within the label layer.

This indicates that improved prediction data, using the best fit attributes would lead to more accurate prediction.

V. CONCLUSION

This study has shown the development and investigation of the effect of the Learning rate using Artificial Neural Network model in the prediction of the results of statistical data.

This study further shows the importance of using appropriate optimization strategies in machine learning to obtain best possible result.

A) Research Highlights

The research highlights of this paper are:

- This paper investigates the effect of the Learning rate on Artificial Neural Network (ANN) in machine learning.

- The approach uses Artificial Neural Network (ANN) and statistical data for implementation.

- The results show various adjusted learning rate with prediction output of 80.65, 80.65, 80.65, 80.65, 83.87, 87.14, 87.1, 87.1, 87.1, 83.87, 80.65, 87.1, 87.1, 90.3, 90.3

REFERENCES

[1] A Orriols-Puig, J. Castillas, and E. Bernado-Mansilla, A comparative study of several genetic-based supervised learning systems. *Studies in Computational Intelligence (SCI). 125*, pp. 205-230, 2008.

[2] F. Akhtar, and C. Hahne, Rapid Miner 5 Operator Reference. *Rapid-I GmbH,* 2012, Retrieved 13:15, February 13, 2015 from: http://rapidminer.com/wp-content/uploads/2013/10/RapidMiner_OperatorReference_en.pdf.

[3] Sengupta, S. (2003). Neural networks and application. Lecture. Department of Electrical and Electronics Engineering. Indian Institute of Technology, Kharagpur. Retrieved from www.youtube.com/watch?v=xbYgKoG4x2g.

[4] Joshua K., (2003). Neural Network Prediction of NFL Football Games. ECE539, Unpublished.

[5] O. U. Anyama, C. P. Igiri, "An Application of Linear Regression & Artificial Neural Network Model in the NFL Result Prediction", International Journal of Engineering Research & Technology (IJERT). Vol. 4 Issue 01, January-2015, 457-461.

[6] C. P. Igiri and E. O. Nwachukwu. "An Improved Prediction System for Football a Match Result". IOSR Journal of Engineering (IOSRJEN), Vol. 04, pp. 12-20, December 2014).

[7] T. Koskela, M. Lehtokangas, J. Saarinen, and K. Kaski. "Time Series Prediction with Multilayer Perceptron, FIR and Elman Neural Networks." Unpublished. Retrieved 14:44, February 13, 2015, from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.35.1631&rep=rep1&type=pdf

[8] C. Stergiou and D. Siganos, "Neural Networks", Sunrise Journal. Vol4, Issue 11 1996. Retrieved 13:01, February 13, 2015 from http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html