International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-5, March 2019

Data Analytics: Role of Activation function In Neural Net

K. Dhana Sree

Abstract: ***Most of the data science contemporary fields like Artificial Intelligence, Machine Learning and Deep Learning are taking advantage of a common model- The Neural Network. Neural network which can learn from experience are becoming popular in solving many of real worlds NP-hard problems. Today any prediction application is taking the support of Neural Network. The accuracy of the neural network models depends on major of the design components like the hidden layers and the activation functions. As we know human brain receives both relevant and irrelevant information at a time and has the capability of segregating both, where the irrelevant can be referred as noise. Just like the human brain the neurons uses activation function to separate the noise from the input and reduce the error. This paper presents the role of hidden layers and activation functions in measuring the accuracy of the Neural Network.***

Keywords: Artificial Intelligence, Machine Learning, Neural Networks, Activation function, Neurons.

I. INTRODUCTION

Trends in data analytics are emerging with a jet speed providing no scope for persistent thoughts. Accurate decision making is an important feature from any corner of data analytics. Supporting this feature Data Analytics is spreading its net for catching any model which can run accurate. Neural network is such a model giving scope of accurate analytics. Neural networks inspired many emerging fields like Artificial Intelligence, Machine learning, Deep Learning and now they are also into cryptography. Machine learning is a computer field which is reflecting its core part in predictive analytics. An unsupervised learning by machines is what all machine learning is. The success of machine learning showed the face of neural network with its well formulated leaning algorithm.

The usage of neural networks to understand deeper relations of data has coined to Deep learning. Deep learning also tasted its success using many folds of the neural networks. Today with their adaptive learning characteristics neural networks are top most models of medical field. The raise and fall of the stock is even analyzed by a neural network. Neural cryptography shows the other corner of Neural Nets for developing data security algorithms. Neural key exchange protocol is widely spread today.

Manuscript published on 30 March 2019.
*****Correspondence Author(s)
K.Dhanasree, Department of CSE, Vardhaman College of Engineering, Hyderabad, Telangana, India

Answering to the question what is making Neural Nets so popular, is the design of the NN. An NN is a network of nodes also called as Neurons. The NN is designed to include various layers like the input, output layers and the most concentrated hidden layer. Between the I/O layers the NN can have hidden layers. The neurons of the I/O layers will be connected to the neurons of the hidden layers. There is no thumb rule on how many hidden layers an NN can have and literature showed the general NNs use two or three hidden layers. Research has proved that hidden layers aim at increasing the predictive accuracy of the NN. The usage of more hidden layers can be seen in deep neural networks, extracting deeper data relationships.

Each layer is connect through of Neurons or nodes of the NN. The input layer neurons take the input values. The output neurons give the predicted output. The hidden neurons which are in the hidden layer play a vital role in prediction functionality. The hidden neurons are defined with an activation function. The activation function is a transfer function, transforming the input to prediction standards. The predictive capability of the neural network is majorly on the type of activation function defined in the hidden neurons.

This paper focuses on the role of hidden layers and the activation function in predictive analytics, their types and the experimental analysis depicting the accuracy. The rest of the paper is organized as follows: section 2 presents the literature survey, more on hidden layers and activation functions is discussed in section 3; section 4 discusses the experimental results of various case studies on hidden layers and activation functions; section 5 concludes the paper.

II. LITERATURE

With the advent of Artificial Intelligence and Machine Learning Neural, networks are becoming popular solving many of the real-world problems. There advances changed the face of machine learning to deep learning. Image processing too is tasting the fruits of Neural Networks [1][2][3]. Being a leading model for many inter-disciplinary fields Neural networks are advancing with more features.

A neural network is a collection of nodes which are similar to Neurons of the human brain [4]. Neural Networks are paving today with their applications on demand in many fields like retail, finance, medical sectors. Neural Networks are basically predictor models and can also be used for classification and image recognitions [5][6].

The NN topology includes the neurons which are interconnected to form various layers. The NN topology includes three different layers the input, output and the hidden layers among which the hidden layer is the concept of concern [7][8] in this work.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

The hidden layer nodes are called as hidden neurons and holding a membership function called as activation function. The activation function of an hidden neuron plays a major role in network performance [9][10]. The activation function is a transformation function, transforming the NN input to a value that will converge to the target output with less error rate [11]. As we know human brain receives both relevant and irrelevant information at a time and has the capability of segregating both, where the irrelevant can be referred as noise. Just like the human brain, the neurons uses activation function to separate the noise from the input and reduce the error. Without the activation function the output of the NN is same as the input.

Research has explored many activation functions: both linear and non-linear [12]. Major of the NN define using nonlinear activation functions as they exhibit non linearity characteristics, taking linear activation functions the input will be same as output without error being minimized. Many non-linear activation functions are in usage. Most popularly used is the sigmoid or the logistic functions which is used for prediction of yes/no cases. The Tanh and the ReLU are other frequently used non-linear activation functions.

III. HIDDEN LAYERS AND NON-LINEAR ACTIVATION FUNCTIONS

The prediction accuracy of an NN is completely based on the usage of number of Hidden layers and type of activation function used. Literature didn't exactly mention on maximum number of hidden layers used, but a thumb rule shows the usage of 2 hidden layers. Literature gave a defined characteristic of NN as- using hidden layers in NN reduces the error.

The predictive power of any NN is defined by the type of activation functions used. Many complex neural networks use non-linear activation functions. Neural network without an activation function defined, is just a linear regression model where the predicted output is same as the given input. This is even same as the NN with linear activation function defined, where the output is same as the input fed with an error. The boundary of linear activation function is linear and the network can adapt to only the linear changes of the input. But most of the real world errors exhibit non-linearity characteristics. Using the linear activation functions the network cannot learn about the erroneous data. Hence most of the NN prefer the usage of non-linear activation functions.

A. Why do we need non-linear activation functions?

The non-linear activation functions are curved boundaries which can adapt to the non-linear changes of the output. The NN model can now learn any complex features of the input that are mapped on to the boundary.

For example let us consider the linear activation function given by

$$Y = F(x) = x \quad (1)$$

Consider the I/O mapping of the linear function given by (1).

Table 1: I/O mapping of Linear function

X	0	1	2	3	4	5	6	10
Y	0	1	2	3	4	5	6	10

Here for the input from 0-10 the boundary of the output is from 0-10. The linear activation function has been trained on the inputs from 0-10. The activation function has adapted to learn the same outputs as that of the inputs. Here the predictive powers of the NN are not so accurate as the ouput is same as input. If the value x=3 has some error then the NN failed to reduce this error as it has given the same output Y=3. This is the major drawback of linear activation functions.

On the other hand consider a non-linear activation function of the sigmoid type given by 2.

$$F(x) = 1/1+e^{-x} \quad (2)$$

Consider the I/O mapping of the sigmoid function.

Table 2: I/O mapping of the non-linear function

X	-E	0	1	2	3
Y	0	0.5	0.7	0.8	0.96
X	4	5	6	E	
Y	0.97	0.98	0.99	1	

Whatever be the input given the output varies between 0 and 1. While the model is training with given data like 0,1,etc the model is adopting to learn the new output like 0,0.5 etc. for a given input boundary the model has learned a new output boundary. This shows the non-linearity characteristic of the sigmoid function. Though the input is linear the output is non-linear. Because of this non-linearity characteristic the sigmoid function is majorly used in the NN.

B. Why non-linear activation functions in NN back propagation algorithm:

This section discusses the mathematical concepts reflecting the neural network back propagation algorithm. These mathematical concepts build a strong knowledge on how exactly an NN works efficiently by defining within itself these concepts.

C. Gradient descent

In general gradient means slope. The slope of a function f is given by its derivative. Gradient descent methods are used to find the local minimum of a function. They are optimization methods, which are used by the back-propagation algorithm to minimize the error. The descent method starts with an initial set of parameters and gradually descents towards the parameters that minimizes the function. The general approach for the gradient descent approach is setting the derivative of a function to be 0 and solve for the parameters. Machine learning uses gradient descent to update the parameters and minimize the error function used in the model.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

D. NN Back propagation

Back propagation is a learning algorithm for a multi-layer Network. The back-propagation algorithm propagates the error through back layers of the network using the gradient descent approach. The algorithm starts with some initial weights and finds the error, which is the difference between the actual target value and the algorithm predicted value. The algorithm finds the derivative of the error function with respect to the network weights. Using the gradient descent approach the algorithm finds the local minimum. At this local minimum the weights are updated and new reduced error is calculated and this error is again propagated back through the layers to update the weights. In the process of back-propagation the networks also learns the weights and trains itself to reduce the error.

E. Learning rate of NN

Learning rate of an NN is how fast the network learners the changes. It is a controlling parameter on how much the weights are adjusted. Small learning rate makes the network to learn very slowly where as fast learning rate makes the network to descent very fast reflecting the chances of missing some of the points unlearnt. What followed by NN is a systematic approach of finding an optimal learning rate. The cycle learning rate method initially sets its learning rate to be very small and systematically adjusts the learning rates in a cyclic way to find the optimal rate.

F. Why non-linear activation functions

Most of real world data is non-linear in nature. Linear data can be perfectly handled as it exhibits usual data patterns with a one to one relationship between the input and the output. Non- linear data patterns are not usual, they continuously gets changed. To model these non-linear data patterns the non-linear activation functions are used in NN models.

The linear functions have a constant derivate and exhibit a constant rate of change, hence they have a constant descent at every iteration. Hence using a linear activation function the NN can learn only on fixed patterns which are common in all iterations. Unusual data characteristics cannot be addressed by the linear activation functions.

With the advent of large data bodies the need for studying the non-linear characteristics of data has grown. Methods which can learn them are focused. Non-linear activation functions came to usage. The non-linear activation functions are curved in nature with many gradients. At these gradients, functional optimal exists. The main design principal of the NN is to minimize the error which is the difference between the actual value and the predicted value. Supporting this constructive principal the NNs use activation functions which can converge to local minimum. On this non-linear boundary of the activation function the NN learns possible unusual data patterns.

IV. ANALYSIS

This section focuses on the role of Hidden layers and non-linear activation functions in NN prediction accuracy. We studied the case of predicting the strength of a brick as predicted by a NN under various activation functions with increased number of Hidden layers.

The building bricks are made using various materials like: cement, slag, ash, water etc. we have taken a data set including 1000 tuples, labeling the strength of the brick with varied material proportions. All the studies are implemented in the R programming environment. The results are studied under varied cases of hidden layers and activation functions.

A. Simulation with Non-linear activation functions

This section presents the simulation results of using non-linear activation functions like Tanh and Sigmoid. The error rate is observed for both.

Case 1: NN with one Hidden layer and Tanh, Sigmod as activation functions.

Figure 1 shows the observed error of NN implemented with one hidden layer. The activation function taken is Tanh. The observed error is 6.912.

Figure 2 shows the observed error of NN implemented with one hidden layer, with Sigmoid activation function. The observed error is 5.668.

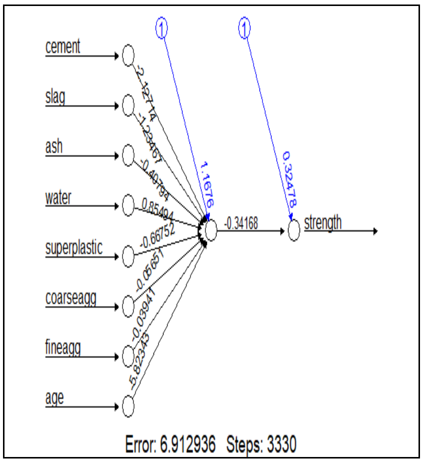

Figure 1: NN: one Hidden layer, Tanh activation function

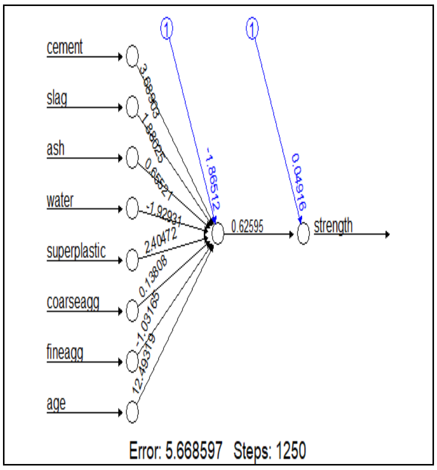

Figure 2: NN: one Hidden layer, Sigmoid activation function

Case 1 shows using Sigmoid activation function the error has descended quickly at less number of steps when compared to that of using Tanh activation function.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Case 2: NN with Two Hidden layers and Sigmod as activation functions.

Figure 3 shows NN with two hidden layers, using sigmoid activation function. Here the observed error is 3.662.

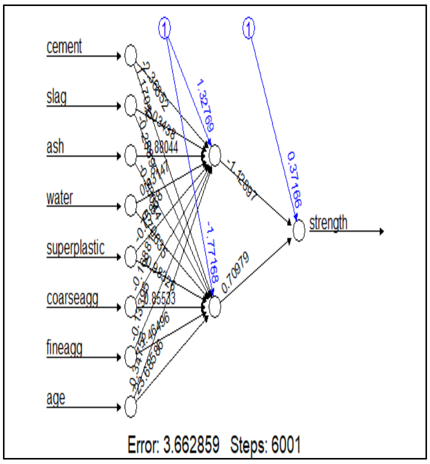

Figure 3: NN: Two Hidden layers, Sigmoid activation function

Case 2 shows on using two hidden layers the error has reduced when compared to usage of one hidden layer.

Case 3: NN with Three Hidden layers and Sigmod as activation functions.

Figure 4 shows the observed error of NN with three hidden layers, with sigmoid activation function is 3.462.

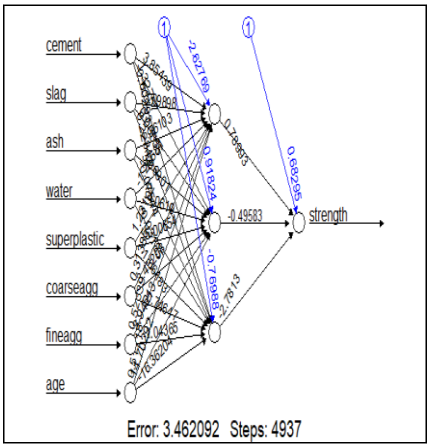

Figure 4: NN: Three Hidden layers, Sigmoid activation function

Case 3 shows the error on using three hidden layers is less when compared to using two hidden layers. The error rate for varied hidden layers is tabulated as shown in Table 3.

Table 3: Error for varied Hidden layers

No.of Hidden Layers	Error rate
1	5.66
2	3.66
3	3.46

V. CONCLUSION

Most of the data science contemporary fields like Artificial Intelligence, Machine Learning and Deep Learning are taking advantage of a common model- The Neural Network. Neural network which can learn from experience are becoming popular in solving many of real worlds NP-hard problems. Today any prediction application is taking the support of Neural Network. The accuracy of the neural network models depends on major of the design components like the hidden layers and the activation functions. Today with their adaptive learning characteristic neural networks are top most models of medical field. The raise and fall of the stock is even analyzed by a neural network. Neural cryptography shows the other corner of Neural Nets for developing data security algorithms. Neural key exchange protocol is widely spread today.

This work presents the experimental observations on implementing these classic models with increased hidden layers and varied activation functions. In general NN with two hidden layers are preferred models.

REFFERENCES

1. M E Petersen,D de Ridder,H Handels,“Imageprocessing with NN: A review”,A Book on Pattern recognition, Elsevier, Vol 35(10),2002.
2. N Plaziac, “Image Interpolation using Neural Networks”, IEEE Transactions on Image Processing, Vol 8(11), 1999.
3. H A Rowley, S Baluja, T Kanade“Neural Network based face detection”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 20(1), 1998.
4. M Y Rafiq, G Buggman,“Neural Network design for engineering applications”, Computers and structures, Vol 79 (17), 2001.
5. B D Replay “Neural Networks and related methods for classification”, Journal of the royal statistical society, Vol 56(3),1994.
6. R P Lippmann“Patten clasiification using Neural Networks”, IEEE communications Magazine, Vol 27(11), 1989.
7. J de Villeirs, E Barnard, “Back propagation Neural Networks with one and two hidden layers”, IEEE Transactions on Neural Networks, Vol 4(1), 1993.
8. K O Stanley, R Mikkulnain“Efficient evolution of Neural Network Topologies”, Proceedings of the 2002 congress on evolutionary computation, 2002.
9. D F Specht, “A general regression neural Network”, IEEE Transactions on Neural Networks,1991.
10. B Warner, M Misra,“Understanding neuralnetworks as statistical tools”, The American Statistician, 1996.
11. V P Plagianakos, M N Vrahatis“Training Neural networks with threshold activation functions and constrained integer weights”, Proceedings of the IEEE INNS IJCNN 2000.
12. E Trentin,“Networks with trainable amplitude of activation functions”, Neural Networks, Elsevier, Vol 14 (4), 2001.

About the Author

Dr. K. Dhanasree received her Ph.D degree in Computer Science and Engineering from JNTU, Anantapur. Her research interests include Data Analytics, Data Mining, Machine Learning, Database Security, and Network Security.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)