International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 08, 2020
ISSN: 1475-7192

Implementation of Artificial Neural Networks using Arcsinh activation function

[1]S. S. Saranya, [2]Ashu Prasad, [3]Ankit Khare

***ABSTRACT**--This paper proposes the use of Arcsinh as an activation function to improve the efficiency of an Artificial Neural Network. The characteristics of sigmoid and Arcsinh have been articulated and Iris Dataset has been used to measure its performance. The paper attempts to strengthen its proposition of using Arcsinh activation function for multi class classification instead of sigmoid activation function by plotting the cost function of each of the activation functions with respect to the number of iterations. The plot obtained thus reveals that the Arcsinh activation function provides a better cost value when compared to the sigmoid activation function.*

***Keywords**--Artificial Neural Networks, ActivationFunctions,Training, Backpropagation*

I. INTRODUCTION

A Neural Network (NN) is computational model which is inspired from the organic sensory systems, to process information. The salient feature of the model is found in its unique architecture of its datahandling framework. It is made up of a large number of nodes that work synchronously to overcome difficult challenges. Artificial Neural Networks, like humans, learn from training. An ANN is specialized for unique tasks, such as segregation of data, medical diagnosisandmarketingvia an iterative training process. Training in the nervous system entails reinforcing certain neuralconnectionswhile weakening the others in the network. ANNs work in a similar way.

Neural networks can be utilized to extricate designs and identify patterns that are otherwise too perplexing to be seen by the naked eye and existing computational models. A trained model is proficient in the field of application it works on. This proficient model provides predictions to unique scenarios and provides answers to different questions. Advantages of this model are:

1. Learning Adaptively: To learn how to do tasks based on the data given for training or initial experience.
2. Self-Sufficiency: A neural network can produce its own rendition of the data it is fed while training and identify patterns of data.
3. Real-Time Environment: Artificial Neural Network is capable of carrying out its processes in a parallel fashion. This has led to several hardware devices being produced to capitalize on this property.
4. Fault Tolerance due to Redundancy: Damage to a network can lead to poor performance. However, several network functions can continue to be in operational capacity.

[1] *Assistant Professor (O.G)SRM Institute of Science and Technology*

[2] *Undergraduate Student, SRM Institute of Science and Technology*

[3] *Undergraduate Student, SRM Institute of Science and Technology*

Received: 21 Dec 2019 | Revised: 18 Jan 2020 | Accepted: 05 Feb 2020

II. ACTIVATION FUNCTIONS

Activation functions are an integral component of a Neural Network to learn about the complex data andconnections connecting the input andoutput nodes. They establish the nonlinearity in the network. The primary task is to form an output signal by processing the inputs from the nodes. That output thus received acts as an input for the successive layer.

In an ANN we work by using inputs (**X**), Weights (**W**) and Bias (**B**). This whole thing can be cumulatively be denoted by (**Z**). Activation function **f(x)** uses this data to generatean output of the concerned layer and apply the same as an input to its successive layer.

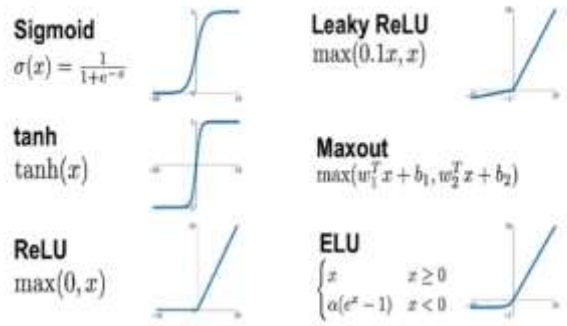
Some popular Activation Functions are:

1. Sigmoid Function (Logistic)
2. Hyperbolic Function - Tanh
3. Rectified Linear Units Function

- ReLU

4. Leaky ReLU Function
5. Maxout Function
6. ELU – Exponential Linear

Units Function



Arcsinh Activation Function: Arcsinh or Inverse Sinh Function is the proposed activation function to be used in the projects. The activation function leads to a better and more accurate prediction of outputs when inputs are provided to the trained neural network.

The Arcsinh function is able to overcome the several problems that are encountered in Sigmoid & Tanh activation functions:

1. Vanishing Gradient – Sigmoid and Tanh functions tend to flatten out as the value of 'x' increases. Arcsinh retains its gradient even with higher values of 'x'.

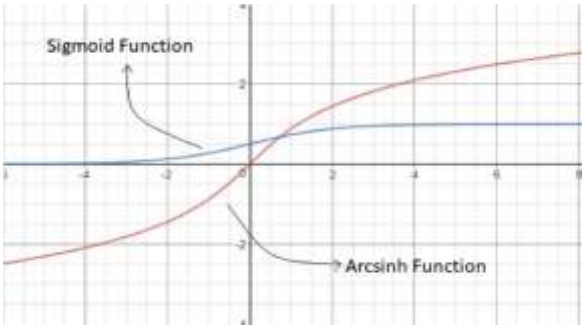
2. Slow Convergence – Sigmoid and Tanh tend to have a slower conversion towards an optimal value of weights.

3. Slower Learning Rate – Due to

slow convergence towards an optimal value, these networks tend to learn slowly.

Received: 21 Dec 2019 | Revised: 18 Jan 2020 | Accepted: 05 Feb 2020

When compared to the Sigmoid Activation Function on the IRIS Dataset, the Arcsinh activation function provided a better and more trained ANN which led to more accurate results. This working is shown in the next topic.



Sigmoid vs Arcsinh

III. TRAINING SUPERVISED TRAINING.

Input and Output labels are already provided in this type of training. The network analyzes the training data and collates its computed output data with the actual output labels. Differences in output are then sent back to the network, tweaking the weights in the neural network. This iterative process continually tweaks the weights of the network up to a satisfactory threshold. The dataset on which the model is trained is known as "training set." During the training phase this data is used multiple times and the weights get tuned.

When we start with our neural network we initialize our weights randomly. Initially the results obtained are not optimal. During the process of training, we start with a poor performing neural network and end up with a network with high accuracy. In terms of the loss function, it is desirable for our loss function to be much lower at the end of training. A neural network improvement is possible, as a change can be achieved by adjusting weights. We want to obtain another function that is superior to the initial one. The aim is to minimize the cost function

A variety of functions exist to optimize cost functions. These algorithms work on both, the information provided by the function as well as the gradients.

Backpropagation is an algorithm used to compute a gradient required in the optimization of weights that is to be used in the network. The term stands for "the backward propagation of errors," because an error is obtained from the output and sent back through the network.

The Delta Rule in machine learning and neural network environments, it is a specific type of backpropagation that helps to perfect connective Machine Learning and Artificial Intelligence networks, making connections between inputs and outputs with layers of artificial neurons. It is used to tweak the weights during backpropagation of an artificial neural network using multiple layers. The delta rule is used in gradient-descent learning.

he Delta rule or the Delta Learning rule is given by:-α is the learning rate (constant)

g(x) is the activation function tjis the target output h_j is the weighted sum of neuron's inpuy$_j$ is the actual output x_i is the input

Received: 21 Dec 2019 | Revised: 18 Jan 2020 | Accepted: 05 Feb 2020

IRIS Dataset

IRIS flower dataset is one of the popular databases for the neural network application. The dataset consists of 3 different types of the flower by implementing pattern recognition. The dataset has three categories of 50 records each, where each category depicts the type of flower. The features that determine each category of the flower has been recorded. The features are discrete, tokenized and continuous in nature. The dataset doesn't have any missing records or non-uniform values and hence doesn't need to be cleaned. The dataset is finalized.

The first category of the flower is distinguishable in a linear fashion from the other two, with the successive two categories not being so. The records of the database being of

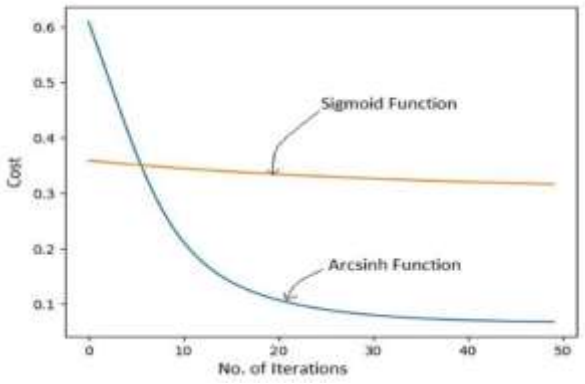
150 tuples in strength have the following features:

- Length of Sepal
- Width of Sepal
- Length of Petal
- Width of Petal

There exists a fifth feature which acts as a label that allows one to distinguish among different categories of the IRIS flower. Those three categories are:

1. IRIS Setosa
2. IRIS Versicolour
3. IRIS Virginica

The cost comparasion of archsinh and sigmoid activation functions was done on the IRIS dataset and the results obtained show that Arcsinh performs better and leads to a more accurate model.



Cost Comparison between Arcsinh and Sigmoid on Iris Dataset

IV. NETWORK ARCHITECTURE

Interconnection can be defined as the method of processing nodes in ANN that are connected to each other. Hence, the arrangements of these

elements and figure of interconnections are a very essential component in the ANN model. These arrangements always have two layers which are common to all network architectures; the input & output layer. The third layer is the Hidden layer which can be composed of multiple layers, in which neurons are neither kept in the input layer norin the output layer. These neurons are hidden from the people who are using the system and acts as an abstraction for the model. On increasing the hidden layers with neurons, the system's processing power can be increased but the training system of the system gets more complex at the same time.

There are five basic types of neuron connection architecture:

- Feed Forward Neural

Network

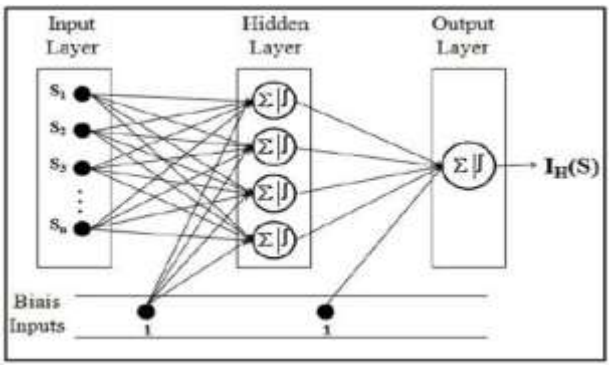
– Single Layered

- Feed Forward Neural

Network

– Multilayered

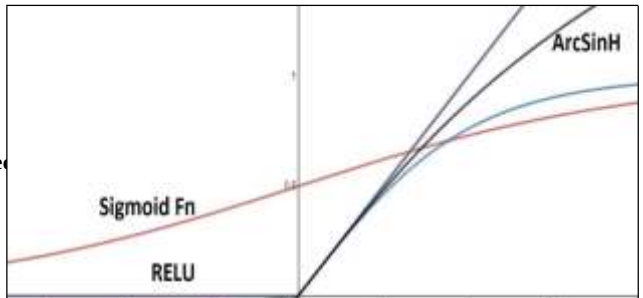
- Feedback with a single node
- Recurrent Network having a single layer
- Multilayer recurrent network



Feed Forward Network – Multilayered

V. POST CONFERENCE COMMENTS & TESTS

More tests comparing the ArcSinH activation function to several other more prominent and prevalent were executed and the results were calculated.



Activation Functions

The tests conclude that ArcSinH activation function achieved a better lower cost model compared to Sigmoid Function but falls off when compared to the more prominent activation functions like RELU and tanH.

Decision of implementing a particular activation function is largely dependent on the type of problem and the results that the model is meant produce. The performance of the activation functions are subject to change if the problem statement itself were to be different. In that regard, it is necessary to be vigilant and choose a suitable activation function, compare performances and select the best models created.

VI. CONCLUSION

The artificial neural network has a lot of potential in the modern technological world. This potential can be leveraged by today's advanced hardware that have immense computational power. They are very flexible and powerful due to their ability to learn via training. Moreover to produce an algorithm to carry out a given task is not needed; i.e. the requirement of understanding the underlying algorithm is eliminated. Their parallel architecture allows such mechanisms to be well suited for real time working environments.

Neural networks contribute to several scientific fields such as medicine, business & finance.

The paper attempts to strengthen its proposition of using Arcsinh activation function for multi class classification instead of sigmoid activation function by plotting the cost function of each of the activation functions with respect to the number of iterations. This leads to a better trained neural network model that can handle larger datasets.

REFERENCES

1. Hudsun, Beale, M., Hagan, M.,& Demuth, H. (2015).Neural Network Toolbox.
2. Xiang Wu, Ran He, Zhenan Sun & Teiniu Tan (2018), A Light CNN for Deep Face Representation With Noisy Labels.
3. Arunselvan Ramaswamy & Shalabh Bhatnagar (2017), Analysis of Gradient Descent Methods With Non diminishing Bounded Errors.
4. Harshit Gupta, Kyong Hwan Jin, Ha Q. Ngyuen, Michael T McCann, Michael Unser (2018), CNN-BPaperedGradientDescent forConsistent CT Image Reconstruction.

Received: 21 Dec 2019 | Revised: 18 Jan 2020 | Accepted: 05 Feb 2020

5. Jun Li, Tong Zhang, Wei Luo, Jian Yang, Xiao-Tang Yuan & Jin Zhang (2017), Sparseness Analysis in the Pre training of Deep Neural Networks.
6. Gonzalo Napoles, Leonardo Concepcion, Rafael Falcon, Rafael Bello & Koen Vanhoof (2017), On the Accuracy– Convergence Tradeoff in Sigmoid Fuzzy Cognitive Maps.
7. Serwa A (2017), Studying the Effect of Activation Function on Classification Accuracy UsingDeepArtificialNeuralNetworks.
8. H.N Mhaskar & C.A Michelli(2016), How to Choose an Activation*Function.*