A Survey on Different Kinds of Activation Functions

S. M. Nahid, Kajol, Abdullah Al Alid, Api adhikary

March 2022

Abstract

Machine learning is one of the most interesting topic in today's time.It excels important and inevitable solutions in various fields.In particular, deep learning is one of the most cost-effective and efficient supervised learning models and it can be applied to various complex problems.Because of its various features it helps us to gain better solutions[1].Neural network is one of the most interesting part in machine learning.

This review discusses the essential attributes of the activation function, and the most widely used activation functions (Sigmoid, tanh, ReLU, LReLU and PReLU). Activation functions and weight initialization methods play an important role in training and operating neural networks[2].Based on the activation function we will determine whether a neuron is activated or not and carries over to the next layer.

1 Introduction

Out of numerous research area machine learning is one of the popular approach as it deals with countless applications[1].Deep learning is a part of machine learning, we can say it is a subset of machine learning. Deep learning is mainly a neural network and the main duties is to actually imitate our human brain and learn from its previous data outcomes. The dazzling growth of data and advances in technology has led to the emergence of various key pathways in deep learning models, making systems more efficient than previous learning models[1].As we already know artificial intelligence more specifically neural networks played key roles to achieve extraordinary outcome. The main intention to construct an artificial neural network so that it can carry out a task, accept different kinds of patterns and grasp this pattern from data[2].

In Section 2 we discussed about previous related research. The discussion and conclusion with valid results have been describe in Section 3 and 4.

A neural network is mainly three parts input layer, hidden layers, and output layer. Inside the hidden layer we have large number of interconnected working units known as perceptrons or neurons.The perceptron consists of four components: an input node, a weight vector, an activation function, and an output

node[2].This activation function is an important concept in neural network-based machine learning.With the help of this we will be able to see the output to the desired area.It is used to determine the output of the neural network, such as yes or no. Displays the resulting values according to the function.There are various types of activation function for example Binary Linear,Sigmoid,Tanh,ReLU,Leaky ReLU,Parameterised ReLU,Exponential Linear Unit,Softmax. In this article, we'll take a look at what's been done recently in this activation function.

2 Literature Review

In this review we are going to use four reference paper associated to different activation functions and related work done so far.As we know that the activation function plays an important role in the decision-making process of the output.

In this paper [2] illustrates some popular activation functions. These are sigmoid, tanh, ReLU, LReLU, PReLU. It states, with the popularity of rectifier nonlinearities usage of sigmoid activation function has declined. It says rectifier nonlinearities, especially ReLU performs good with He normal initialization in particular kinds of networks. It also demonstrates that in Xavier initialization tanh function is used in case of not deeper networks. For deeper networks, normal initialization along with ReLU gets more preference.

In [4] describes neural networks with complex values. Here, includes a common activation function, learning paradigm, inputs and outputs, representation and application.According to this paper complex-valued neural networks (CVNN) are ANNs that process information using complex-valued parameters and variables. CVNN is equivalent to a two-dimensional real-valued neural network. They need activation functions of complex valued neural networks because activation functions introduce non-linearity.In this paper they mainly discuss different wireless communications and many other fields where complex numbers happen naturally or on purpose and neural networks are used.

In this study [1]they focus more on neural network architectures with deep learning they demonstrates design methods, deep networks are analyzed and categorized into generative architectures, discriminate architectures and hybrid architectures. The future research work could be introducing hybrid architectures in convolution neural network for better performance improvement.

On the other hand if we see in A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects [3] here describes various aspects of convolutional neural networks. It includes common building blocks, classic networks, related functions, applications, and prospects. In this paper they introduce activation functions, loss functions and optimizers for CNN. They offered some thumb rules for selecting these functions. According to this paper if the activation function is not used rather than a linear function is used, the hidden layer will have no effect. With non-linear activation function abilities of fitting data enhance in neural networks. Their demonstration starts with a sigmoid function. Basically, the sigmoid function turns a real number into 0 or 1. So, whenever

we need binary classification, the sigmoid function can be used. Furthermore, for SENet and MobileNet v3 the output needs to be converted between 0 and 1. So, In this scenario sigmoid is a good function to implement. Besides, the tanh function turns the output into -1 or 1 for next layer input. So, it helps the next layer learn easily. The ReLU function takes input and if the input is less than 0 the output will be 0, and if not the output will be the input itself. Basically, for it's simplicity it speeds up learning. So, It can be used for deeper networks; it takes less neurons. There are also some variants of ReLU. They are LReLU, PReLU, ELU. For LReLU, output of negative value is not zero rather it is input divided by a. Where a is a fixed parameter ranging between 1 and +infinity. Unlike LReLU the PRe PReLU's negative input's output is not based upon a predefined data. And ELU is another improved version of ReLU. It's negative part is a curve. For this, lots of complicated derivatives are demanded.

3 Discussion

Due to the advantages of activation function in neural networks such as local they are widely used in both research and industrial projects.[3].Neural networks are essentially very powerful machine learning engines that mimic the learning of the human brain.Depending on the complexity of the task, multiple neurons form a complex network to transmit information to each other.We know that a neural network made up of interconnected neurons. Each neuron is characterized by weights, biases, and activation functions.The choice depends on the type of prediction or target the model made using the activation function.In different classification problem we can use different type of neural network.Depending on the type of prediction problem we want to solve, we should choose an activation function for our output layer and what we are going to predicted.

4 Conclusion

This literature review provides an understanding of activation function.Based on the applications and the design methods,we will choose different kind of activation function for our model.If we observed the findings of the survey and summarized which will help everyone in future to improve the research work through any of the models and provide convenient info about basic idea on neural network.

References

[1] Survey on Neural Network Architectures with Deep Learning.

[2] A Survey on Activation Functions and their relation with Xavier and He Normal Initialization

[3] A Survey of Convolutional Neural Networks:Analysis, Applications, and Prospects.

[4] A Survey of Complex-Valued Neural Networks

5 Contribution

Student id & name	**Section No**	**Section Title and Others**
ID: 19-39998-1 Name: S.M.Nahid	[1][2]	Finding all the paper, Abstract , Introduction, Formatting in LaTeX.
ID: 19-40018-1 Name: Kajol	[2],[4]	Read References paper[1],[4], Conclusion
ID: 19-39538-1 Name: Api adhikary	[3],[4]	Help in Conclusion and Discussion Part
ID: 19-40096-1 Name: Abdullah Al Alid	[2],[3]	Read References paper[2],[3], Discussion

Table 1: Section(s) Written in the paper by the group member