International Journal of Engineering Sciences & Research Technology

(A Peer Reviewed Online Journal)
Impact Factor: 5.164

IJESRT

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

IJESRT
INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

THE TWO NOVEL ACTIVATION FUNCTIONS FOR ARTIFICIAL NEURAL NETWORK (ANN)

Renu Bahuguna[*1], Shivankur Thapliyal[2]
[*] HOD-CSEDepartment of Computer Science and Engineering, Doon Group of Institutions Rishikesh (Uttarakhand), India

DOI: https://doi.org/10.29121/ijesrt.v10.i11.2021.3

ABSTRACT

The Whole Exploration and Expansion nature of an Artificial Neural Network is dependent on some Activations functions, such as a Sigmoid function ,Rectified Linear Unit (Relu) , but still they contains some possible exponentially errorneous nature, which we see when we trained an Artificial Neural Network using Multilayer Perceptron contains a large number of layers , so we developed a novel Activations functions which works on the exponentially growth of these layers and when the number of hidden layers would have been increases we will reached a better accuracy and finding a better optimized threshold. We generally discovered two novel Activation Functions, both react with better accuracy as compared to preciously discovered Activation Functions under the space of points. A Negative and Positive number scale represents how we reached a better optimization threshold with better accuracy when we increases the marginal distances between two numbers , the curve also adopts a very smoothly nature which shows the exponential rate and the execution time is more better than some other previously Activations Functions. The growth rate per sample or the growth rate per points would be also improve. To know about detailed of these novel Activation Functions, and explore these Activations Functions, adopt some mathematical background to prove its authenticity.

KEYWORDS: Activation Functions for ANN, Artificial Neural Network (ANN).

1. INTRODUCTION

In the current trends of recent information's age intelligent systems [1] and smart devices [2] play's a major and versatile role to perform typical computational tasks [3] in a very easy and reliable manner with redirect to optimality or In other words with respect to reliable optimization techniques [4,5]. In this intelligent systems [1] and smart devices [2] Artificial Neural Network (ANN) [6, 7, 8] play's a very significant role to perform any complex and typical tasks or selection some features and put on the weight factor and do summation of features and weight and will give to optimization functions [9], which optimize any value in the form of 0s and 1s and redirect to optimality. In the concept of ANN [6, 7, 8] there are various hidden layers which are responsible to perform any desired task with respect to optimizations [10]. In the first layer there are various neurons which are generally used to select any feature and In the concept of hidden layer, we will multiply the desired feature and correspondent weight and we will apply this concepts with all incoming edges on any neurons and will do summation of all theses and pass it to optimization functions [9], which will optimize any values in the form of 0s and 1s. Generally we used Sigmoid [11] and Relu function [12] for optimization or generally called the activation function, but the main drawback of Sigmoid function [11] is that it's also optimized values in the form of 0's and 1's but the derivative of this function lies between 0 to 0.25 and this will occur some major problems such as 'Vanishing Gradient Problem [12, 15]' and 'Stochastic Gradient Problem [13, 14]'. So In this paper we introduce two new optimization functions [9] or activation functions [9] instead of Sigmoid [11] and Relu [12], which also converts any values between the range of 0 and 1 but this respective derivative of sigmoid functions [11] don't lies between 0 and 0.25, it's also lies between 0 and 1 and overcome the major problem of 'Vanishing Gradient Problem [12, 15]' and 'Stochastic Gradient Problem [13, 14]', which are the most reliable features of that. Here In this paper we proposed 4 optimization functions [9] which gives much better results as compared to as Sigmoid [11] and Relu Activation Functions [12]. Generally these optimization functions [9] also called

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

activation functions [9] which are responsible to take decisions such as this values are absolutely occurred or not?, because Sigmoid [11] and Relu functions [12] also contains much advantageous features but it's also contains some drawbacks such as their respective derivative also transforms with much smaller values such as the derivative of sigmoid much lies between the range of 0 to 0.25 and this is the main reason of 'Vanishing Gradient Problem [12, 15]' and 'Stochastic Gradient Problem [13, 14]'. But these newly proposed two activations functions [9] also overcome these major problems and will give a very satisfactory results and these all two newly proposed activations functions [9] also examine with various data sets and observed their desired behaviour. So the detailed concepts and the working mechanisms of these two activations functions [9] are described in the upcoming sections.

2. THE TWO NOVEL ACTIVATIONS FUNCTIONS FOR ANN

In the current trends of present Information age the techniques to manufacture any Intelligent Systems and Smart Devices are one of the most crucial and challenging factor. To creation of any Intelligent Systems and Smart devices Artificial Intelligence play's a major significant and versatile role and the major concepts of Artificial Intelligence such as 'Artificial Neural Network' are responsible to train any systems like as human brain and will also responsible for decision making capability and to perform various complex and rational computational tasks. In the concepts of Artificial Neural Network (ANN) optimization functions or generally called the activations functions play's a major significant role, because the whole mechanisms of Artificial Neural Network (ANN) are totally dependent on the optimization functions or activations functions which are responsible to map any values between 0 and 1, but there are also various optimization functions or activation functions are available for any ANN, but the one most drawback of these all previously developed ANN are that their correspondent derivative also lies between 0 to 0.25 and that's why it's occurred some major problems such as 'Vanishing Gradient Problem' and 'Stochastic Gradient Problem', In this paper we proposed two new optimization functions or activation functions which are responsible to overcome these some major problems and one of the most significant advantageous features of these two newly proposed activations functions are that their respective derivative also lies between the range of 0 and 1 and that's why it's also overcome these major problem of ANN which will frequently occurred with some previously developed optimization functions or activation functions such as Sigmoid and Relu. In this paper we also demonstrate or visualize some graphical representation of these newly proposed two optimization functions or activations functions and also theses these functions with some other real data set and compare it's values with some other previously developed optimization functions or activations functions and it's will give more better results as compared to others. The detailed description of these all two newly proposed optimization functions or generally called activations functions are in the upcoming sections of this paper.

We have some potential previously developed Activations functions like Identity function , Binary Step Function, Bipolar Step Function , Sigmoid function , Binary Sigmoid function , Bipolar Sigmoid Function , Ramp Function but these all functions do a very little bit changes when the marginal length or the marginal weight between two points reached a very peek threshold or In other words we say that when the marginal length between two points become very high or we do a very high change on numbers or a very high jumping then these functions also gave satisfactory results but the values of that function do a very lesser changes in original functional value , Meanwhile if we take some large numbers which is either +ve or –ve or when we put these values on that functions then the finalize or resultant value of these function do some changes at only decimal places and we reached a peek threshold after some large infinity points of numbers, and that's the reason we also reached a better accuracy but we will not find a best optimized improved satisfactory results , and it's also flaws the rate of accuracy and interpretation time or execution time , so this remarkable breakthrough we improvised using a novel Activation function.

Mathematical Analysis

The first Novel Activation function analysis are as follows:

Function : $\boldsymbol{f(x) = \tanh\left(\frac{e^x}{1+e^{-x}}\right)}$ (1)

Where x ranges would be $-\infty \leq x \leq +\infty$

The functional values of this function are as follows, and compared to it to sigmoid function values:

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

Table 1: Value comparison between two novel functions

$f(x) = \tanh\left(\frac{e^x}{1 + e^{-x}}\right)$			$f(x) = \frac{1}{1 + e^{-x}}$	
X	F(x)		x	F(x)
-∞	0		-∞	0
-10	$2.06106*10^{-9}$		-10	$4.539787*10^{-5}$
-9	$1.52281*10^{-8}$		-9	$1.233946*10^{-4}$
-8	$1.124974*10^{-7}$		-8	$3.353501*10^{-4}$
-7	$8.307712*10^{-7}$		-7	$9.110512*10^{-4}$
-6	$6.12902*10^{-6}$		-6	0.0024726232
-5	$4.509607*10^{-5}$		-5	0.0066928509
-4	$3.294289*10^{-4}$		-4	0.01798621
-3	0.0023611908	**Compared**	-3	0.047425873
-2	0.016130962		-2	0.11920292
-1	0.098616453		-1	0.26894142
0	0.46211716		0	0.5
1	0.96311369		1	0.73105858
2	0.99999555		2	0.88079708
3	1		3	0.95257413
4	1		4	0.98201379
5	1		5	0.99330715
6	1		6	0.99752738
7	1		7	0.99908895
8	1		8	0.99966465
9	1		9	0.99987661
10	1		10	0.9999546
+∞	1		+∞	1

We easily see that in a table for the function $(x) = \tanh\left(\frac{e^x}{1+e^{-x}}\right)$, when the number of numeric identity marginal distance would be increases , there is a major changes in the resultant value , and we easily achieve an appropriate value 1 and this indicates that the constantable growth of this function is outstanding and that's why we reached a threshold point of accuracy.

Now for the function $f(x) = \frac{1}{1+e^{-x}}$ we easily see that in the table , that when the number of numeric identity or marginal distances would have been increases there are very small changes at only decimal side , and we get a contant rate of growth of this function after many values fluctuations , so that's why some exponential error still existing in sigmoid function , which create some flaws at that time when we train an Artificial Neural Network along with Multi Layer Perceptron , whereas there are a huge number of hidden layers, but our function optimize that situation and gives a better optimized threshold value at very early stages.

Graph Representation of function $f(x) = \tanh\left(\frac{e^x}{1+e^{-x}}\right)$

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

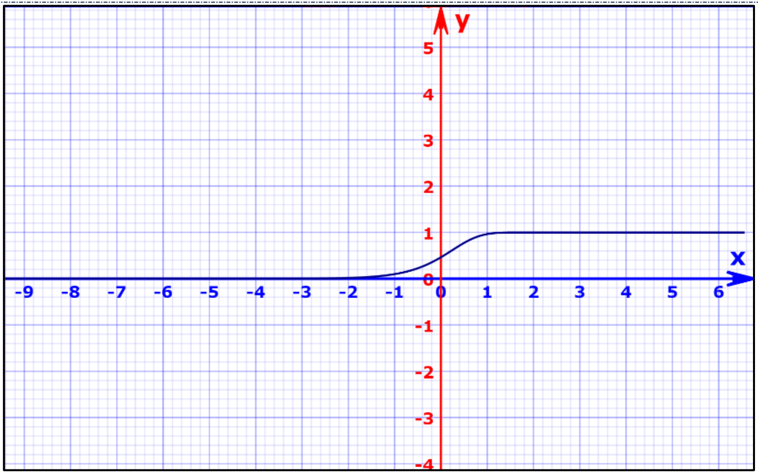


fig1: very high marginal distances

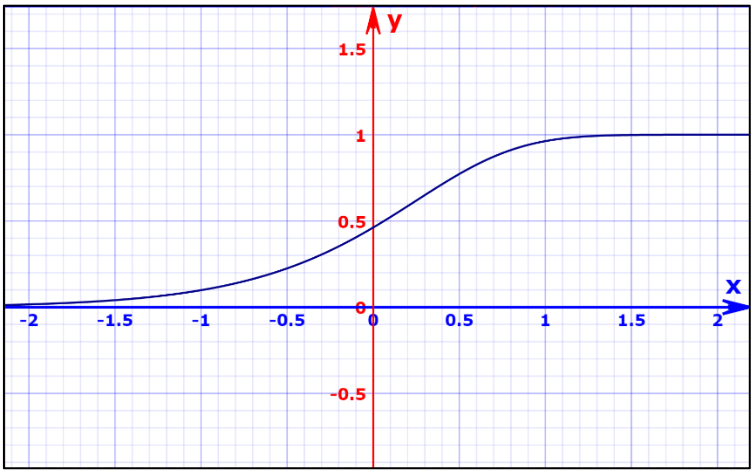


Fig2: Medium marginal distances

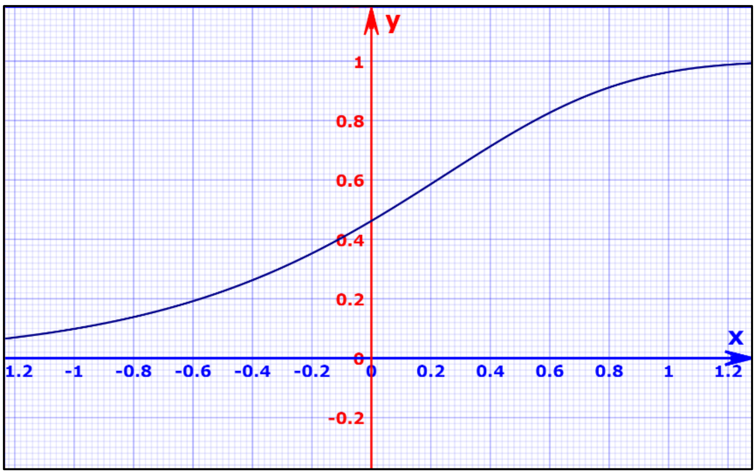


Fig3: small marginal distances

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

Now we take a very small marginal distances to see the variations :

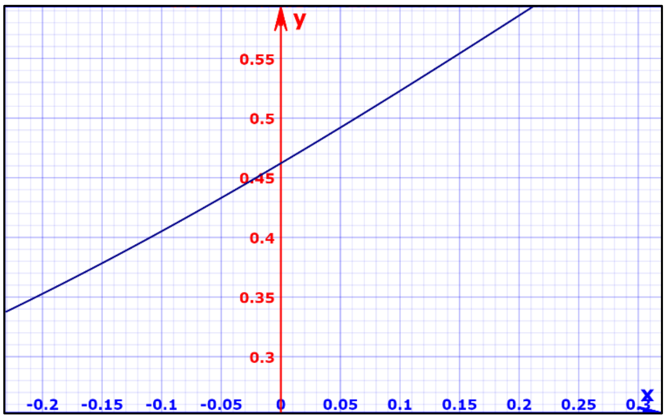


Fig 4: very small marginal distances

Now you easily see that when we reached a very small marginal distances from high to low then you see the exact variations between them.

Now Comparison this function to sigmoid function and see the actual variations :

$$f(x) = \tanh\left(\frac{e^x}{1+e^{-x}}\right) \qquad\qquad f(x) = \frac{1}{1+e^{-x}}$$

Comparison Graph

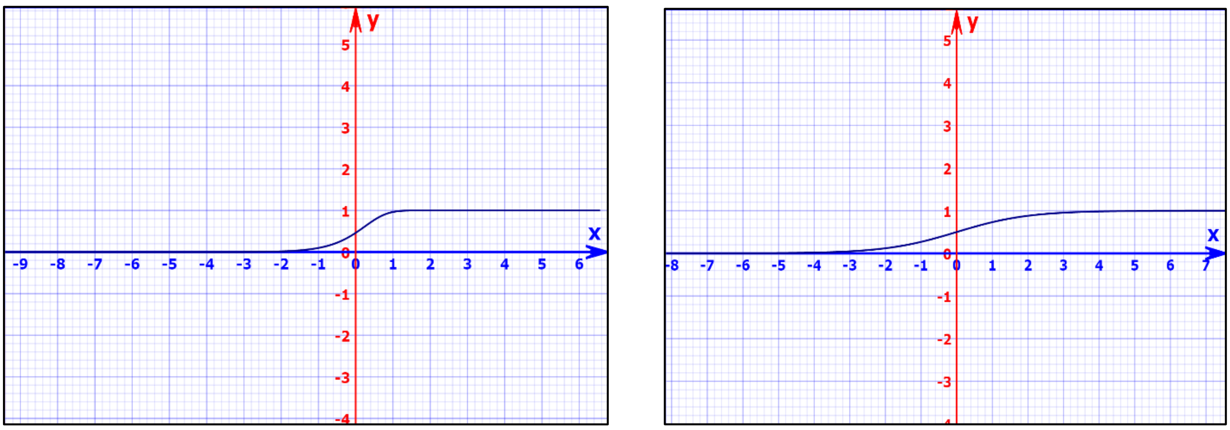


fig1: very high marginal distances

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

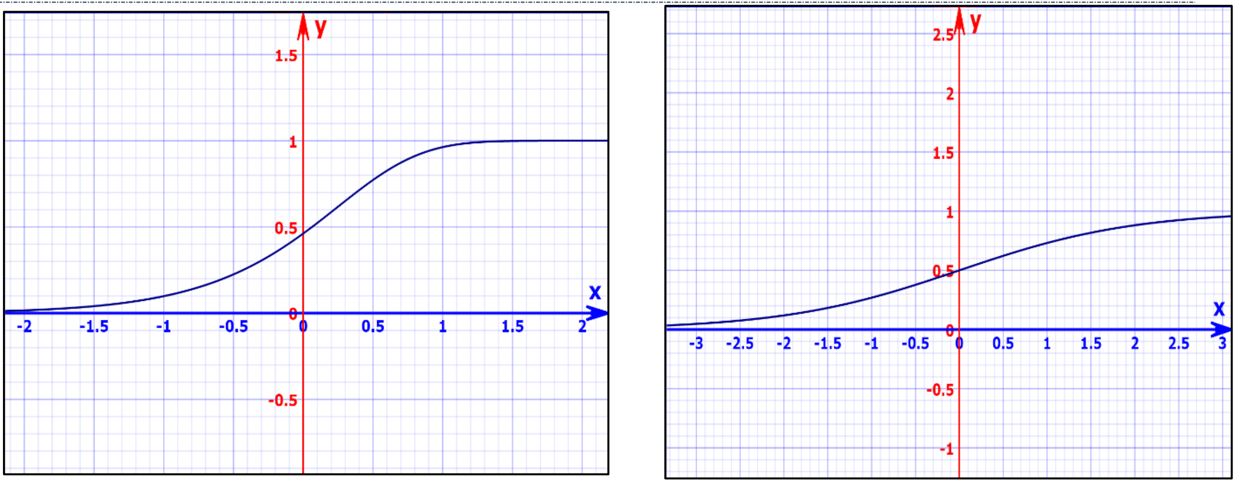


Fig2: Medium marginal distances

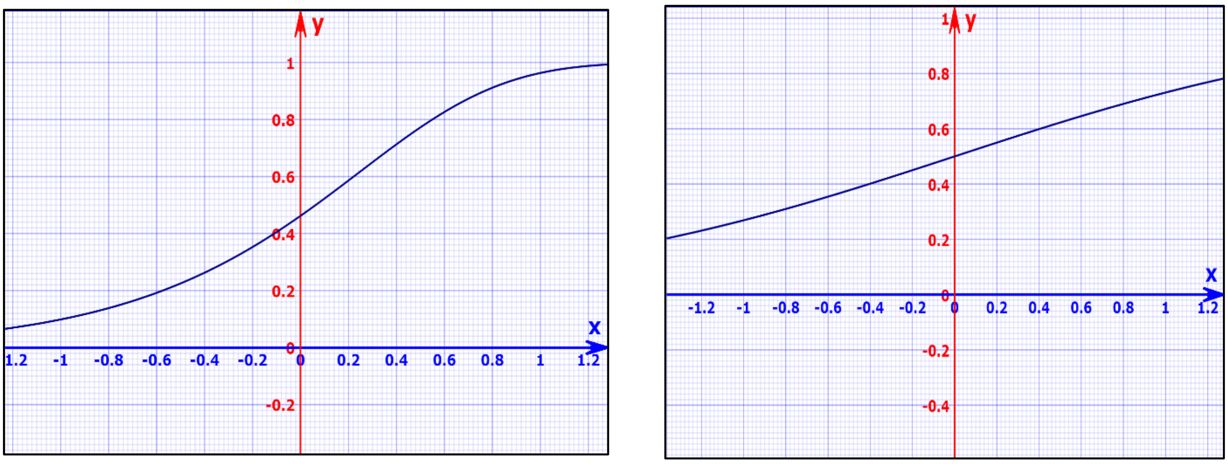


Fig3: small marginal distances

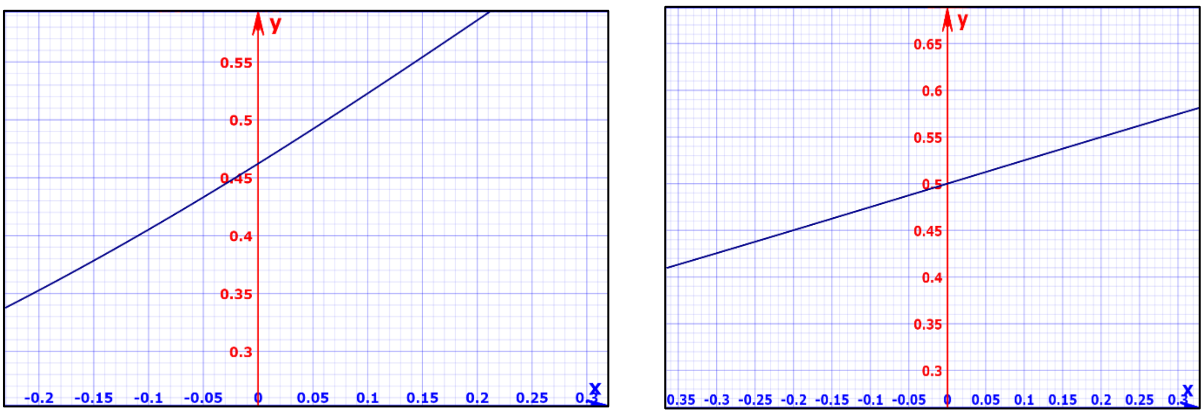


Fig 4: very small marginal distances

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

Now you easily see that how we reached a global optimize threshold value at very early stages using this $f(x) = \tanh\left(\frac{e^x}{1+e^{-x}}\right)$ function

Now Second Novel function is :

$$f(x) = \tanh\left(\frac{1}{e^{-x}}\right) \qquad (2)$$

Where x ranges would be $-\infty \leq x \leq +\infty$

The functional values of this function are as follows , and compared to it to sigmoid function values:

Table 2: Comparison graph of this function to sigmoid function

$f(x) = \tanh\left(\frac{1}{e^{-x}}\right)$			$f(x) = \frac{1}{1+e^{-x}}$	
x	F(x)		x	F(x)
-∞	0		-∞	0
-10	$4.539993*10^{-5}$		-10	$4.539787*10^{-5}$
-9	$1.234098*10^{-4}$		-9	$1.233946*10^{-4}$
-8	$3.354626*10^{-4}$		-8	$3.353501*10^{-4}$
-7	$9.118817*10^{-4}$		-7	$9.110512*10^{-4}$
-6	0.0024787471		-6	0.0024726232
-5	0.006737845		-5	0.0066928509
-4	0.018313591		-4	0.01798621
-3	0.049745973	**Compared**	-3	0.047425873
-2	0.13451504		-2	0.11920292
-1	0.35213549		-1	0.26894142
0	0.76159416		0	0.5
1	0.99132892		1	0.73105858
2	0.99999924		2	0.88079708
3	1		3	0.95257413
4	1		4	0.98201379
5	1		5	0.99330715
6	1		6	0.99752738
7	1		7	0.99908895
8	1		8	0.99966465
9	1		9	0.99987661
10	1		10	0.9999546
+∞	1		+∞	1

We easily see that in a table for the function $f(x) = \tanh\left(\frac{1}{e^{-x}}\right)$, when the number of numeric identity marginal distance would be increases , there is a major changes in the resultant value , and we easily achieve an appropriate value 1 and this indicates that the constantable growth of this function is outstanding and that's why we reached a threshold point of accuracy.

Now for the function $f(x) = \frac{1}{1+e^{-x}}$ we easily see that in the table , that when the number of numeric identity or marginal distances would have been increases there are very small changes at only decimal side , and we get a contant rate of growth of this function after many values fluctuations , so that's why some exponential error still existing in sigmoid function , which create some flaws at that time when we train an Artificial Neural Network along with Multi Layer Perceptron , whereas there are a huge number of hidden layers, but our function optimize that situation and gives a better optimized threshold value at very early stages.

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

Graph Representation of function $f(x) = \tanh\left(\frac{1}{e^{-x}}\right)$

$$f(x) = \tanh\left(\frac{1}{e^{-x}}\right) \qquad\qquad f(x) = \frac{1}{1+e^{-x}}$$

Comparison Graph

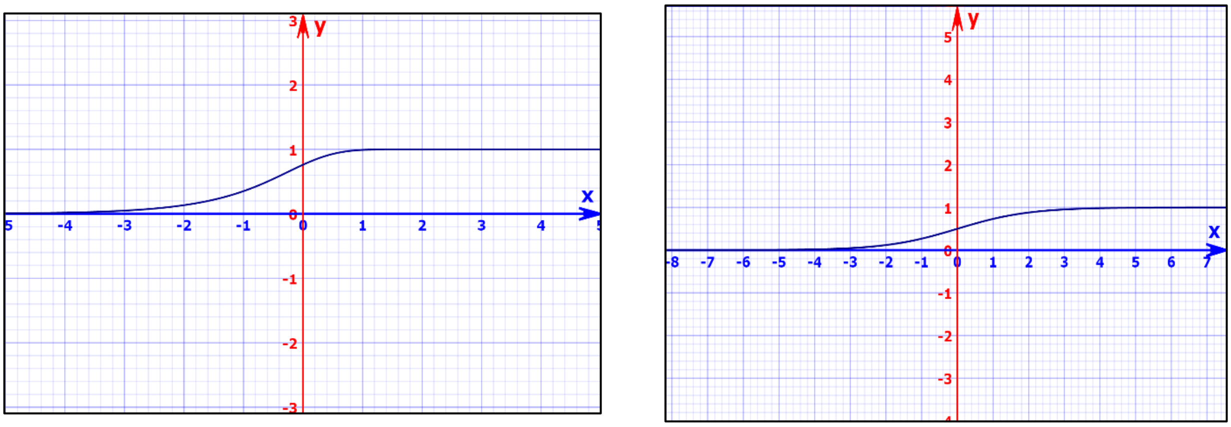


fig1: very high marginal distances

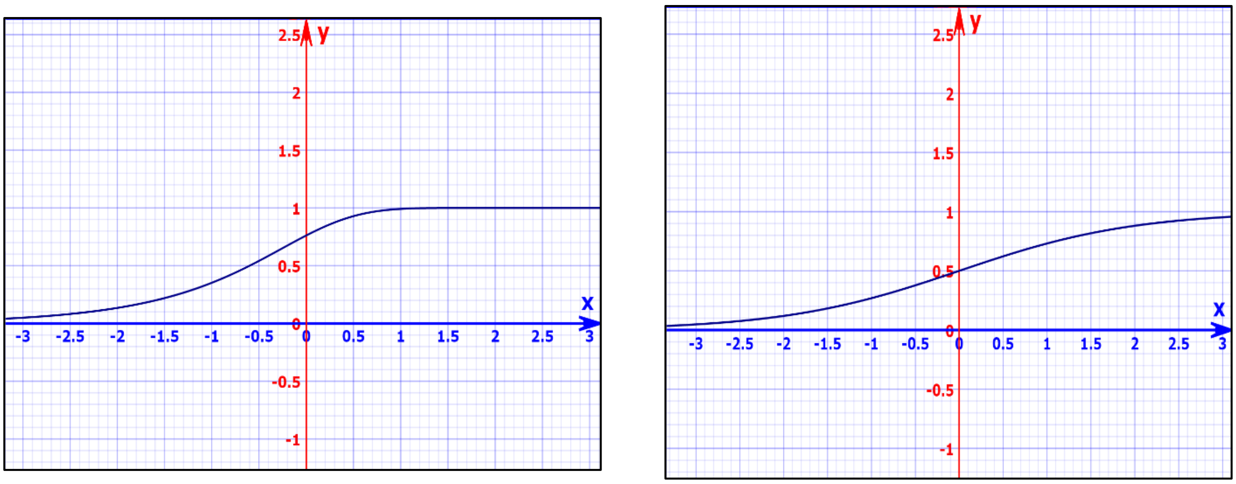


Fig2: Medium marginal distances

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

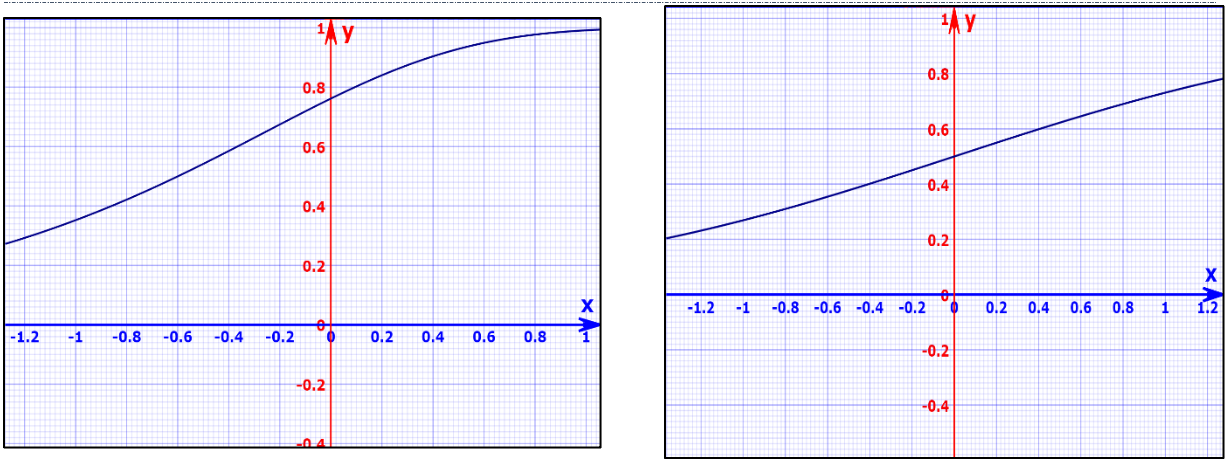


Fig3: small marginal distances

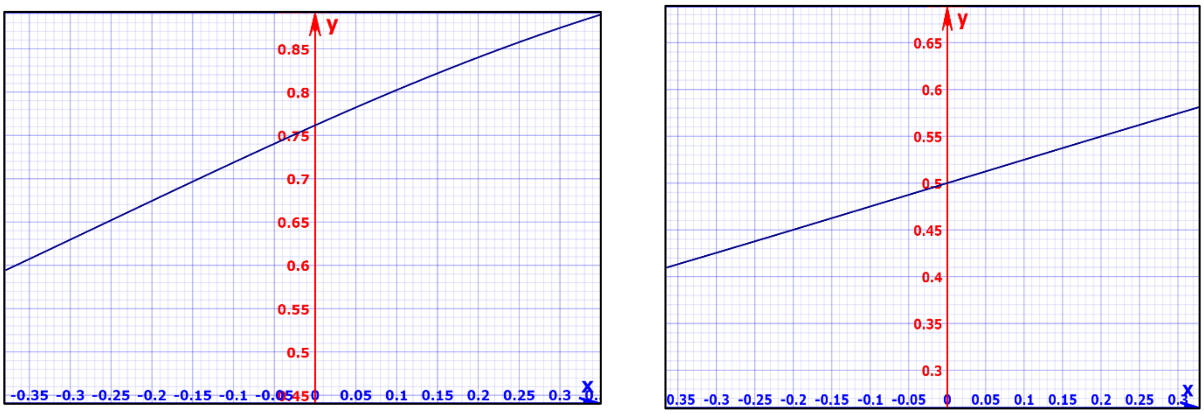
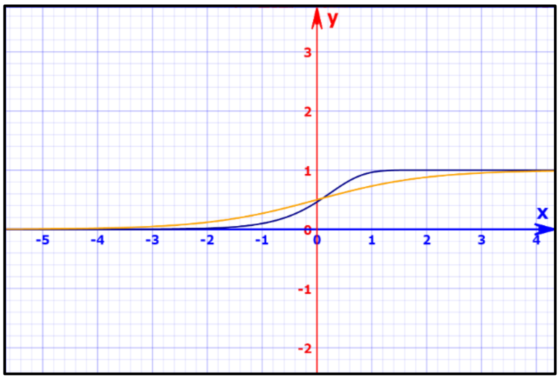


Fig 4: very small marginal distances

You easily see that with this graph that how we reached a constant rate of growth in very early stages using this $f(x) = \tanh\left(\frac{1}{e^{-x}}\right)$ function.

Comparison graph of function $f(x) = \tanh\left(\frac{e^x}{1+e^{-x}}\right)$ and sigmoid function $f(x) = \frac{1}{1+e^{-x}}$

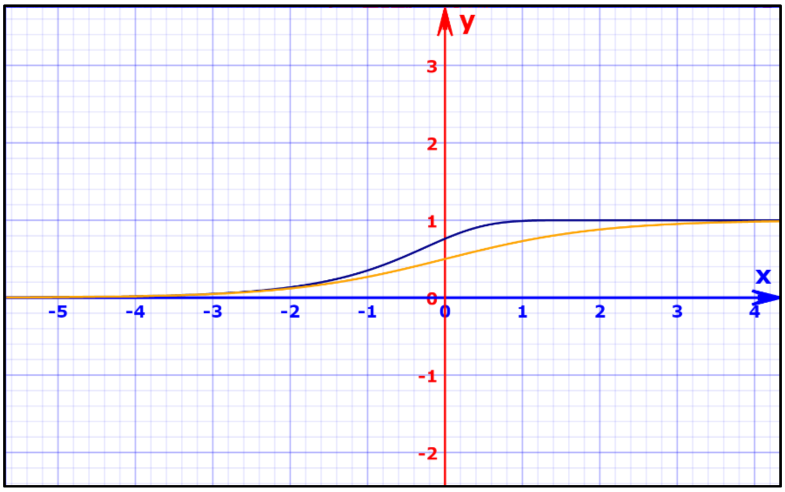


Where Blue line represents function: $f(x) = \tanh\left(\frac{e^x}{1+e^{-x}}\right)$

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

And yellow line represents function: $f(x) = \frac{1}{1+e^{-x}}$

Comparison graph of function $f(x) = \tanh\left(\frac{1}{e^{-x}}\right)$ and sigmoid function $f(x) = \frac{1}{1+e^{-x}}$



Where Blue line represents function: $f(x) = \tanh\left(\frac{1}{e^{-x}}\right)$
And yellow line represents function: $f(x) = \frac{1}{1+e^{-x}}$

3. RESULTS AND DISCUSSION

Data set	Function	Classification rate	Function	Classification rate
XOR dataset	$f(x) = \frac{1}{1+e^{-x}}$	100	$f(x) = \tanh\left(\frac{e^{x}}{1+e^{-x}}\right)$	100
Baloon dataset		100		100
Iris dataset		92		94
Cancer dataset		98		98
Heart dataset		87.5		88

Data set	Function	Classification rate	Function	Classification rate
XOR dataset	$f(x) = \frac{1}{1+e^{-x}}$	100	$f(x) = \tanh\left(\frac{1}{e^{-x}}\right)$	100
Baloon dataset		100		100
Iris dataset		92		95.3333

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

Cancer dataset		98		98
Heart dataset		87.5		92

4. FUTURISTIC SCOPE

In the present scenario of today's information age there are various tasks of computing which required a lot of computations as well as redirect to optimality. So how we make possible that we applying perfect optimization techniques with these typical computational tasks and redirect to optimality, these scenario are totally dependent on the optimization functions and In generally activation functions in the reference of Artificial Neural Network (ANN). So here In this paper we proposed two optimization functions which contains a dynamic nature and optimized structure which also mapped any values between the range of 0's and 1's. In previous scenario there are also available much activations functions which increases the reliability and optimality of Artificial Neural Network (ANN), but these activations functions also contains some much of disadvantages such as these functions also mapped value between 0's and 1's but these functions derivatives which are responsible to update weight during 'Back-Propagation' also lies between 0 to 0.25, this is the main disadvantage of these previously developed activations functions and with this reason various problems such as 'Vanishing Gradient Problem' and 'Stochastic Gradient Problem' also will be occurred. In these problem the new updated weight becomes more smaller and many times it's equal to old weight and doesn't redirect to Global Minima, and many time these weights becomes more larger and also doesn't redirect to Global Minima. So with this newly proposed two activations functions we also overcome to this problem. Because the derivative of these newly proposed two activations functions doesn't lies between 0 to 0.25 rather as Sigmoid functions, it's also lies between 0's and 1's and will overcome the problem of 'Vanishing Gradient Problem' and 'Stochastic Gradient Problem'. In the near future we also improvised these newly proposed two activations functions and also will optimize these functions for better and improvised results observation. These new activations functions also tested with various data sets and also observed and compare their desired results with some previously developed some activations functions such as sigmoid and Relu activation functions. Because these all major problems such as 'Vanishing Gradient Problem' and 'Stochastic Gradient Problem' will occur during the activity of Back Propagation because weights would updated during Back Propagation. So In the near future there are various advantageous scope for this research because optimization is the main key feature of any Artificial Neural Network and, we will also discovered some more activations functions to improvised these previously and newly proposed activations functions. Because the total decision making capability and the working mechanisms of any Artificial Neural Network (ANN) are totally dependent on the working mechanisms of these optimization functions or activations functions. So it's to be essential that we improvised and optimized any previously developed optimization functions or activation functions to make Artificial Neural Network more reliable and superior for any real time system based computational tasks. So Various advancements have to be done in the near future with these optimization function, In this paper some newly proposed these 4 activation functions are fully responsible to optimized or mapped any value between 0 and 1 and also it's derivative lies between the range of 0 to 1 rather 0 to 0.25 as Sigmoid and the major advantageous features is that this two activations functions also overcome some significant problem of ANN such as 'Vanishing Gradient Problem' and 'Stochastic Gradient Problem'.

5. CONCLUSION

In the concept of Artificial Neural Network (ANN) optimization functions or generally called the activations functions play's a major and significant role to do various typical and complex computational tasks. But the reliability and accuracy of these Artificial Neural Network (ANN) are totally dependent on some optimization functions or some activation functions. Because when Back Propagation will applied to update correspondent weights for redirection to optimality and optimize the Loss factor of actual output and the predicted output so these optimization or activations functions play's a major role. In this paper we improvised these previously developed optimization or generally called the activations functions and will developed newly proposed two optimization functions or activations functions. These newly proposed two optimization functions will gives a much better results some previously developed some optimization or activations functions. Because these functions will check some real time data sets and examine or compare their current results or behavior some previously developed some optimization functions or activation functions. In this paper the major achievement is

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

that In the concept of sigmoid functions it's also mapped any values between 0 and 1 but their derivative lies between the range of 0 to 0.25 and this will occur some major problems such as 'Vanishing Gradient Problem' and 'Stochastic Gradient Problem', but using these two newly improvised optimization functions or activations functions we also overcome these major or significant problems of Artificial Neural Network (ANN) because it's derivative lies between the range of 0 and 1 rather 0 to 0.25 as sigmoid, so that's why it's overcome some major problems of ANN and also provide a better reliability to perform any complex and typical computational tasks. Here In this paper we represents some graphical presentation of previously developed optimization or activation functions or newly proposed optimization functions or activations functions.

6. ACKNOWLEDGEMENT

We are very much grateful to all respected professors of DIET Rishikesh for their kind help, lasting encouragement, valuable suggestion throughout the entire period of our project work. We are highly indebted to his astute guidance, sincere support and boosting confidence to make this Research successful.The acknowledgment will be incomplete if we fail to express our obligation and reverence to our family members and friends whose moral support is great factor in doing this research.

7. FUNDING

This study was not funded by any other profitable or non-profitable Organizations.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

[1] Haykin S., "Neural Networks A Comprehensive Foundation", 2nd edition, Pearson Education, 19a99.
[2] Ms. Sonali. B. Maind et. al., Research Paper on Basic of Artificial Neural Network, International Journal on Recent and Innovation Trends in Computing and Communication Volume: 2 Issue: 1 | January 2014.
[3] Vidushi et al., International Journal of Advanced Research in Computer Science and Software Engineering 2 (10), October2012, pp.278-284.
[4] About Feed Back Network from website http://www.idsia.ch/ ~juergen/rnn.html .
[5] Sucharita Gopal, "Artificial Neural Networks for Spatial Data Analysis", Boston, 1988.
[6] Eldon Y. Li, "Artificial Neural Networks and their Business Applications", Taiwan, 1994. FLEXChip Signal Processor (MC68175/D), Motorola, 1996.
[7] Christos Stergiou and Dimitrios Siganos, "Neural Networks".
[8] About Neural Network from website http://en.wikipedia.org / wiki/Neural_network .
[9] Girish Kumar Jha, "Artificial Neural Network and its Applications", IARI New Delhi.
[10] Image of a Neuron form website http://transductions.net/ 2010/02/04/313/neurons/.
[11] Ugur HALICI, " Artificial Neural Networks", Chapter 1, ANKARA.
[12] Ajith Abraham, "Artificial Neural Networks", Stillwater,OK, USA, 2005.
[13] Lippmann, R.P., 1987. An introduction to computing with neural nets. IEEE Accost. Speech Signal Process. Mag., April: 4-22.
[14] Carlos Gershenson, "Artificial Neural Networks for Beginners", United kingdom.
[15] Concepts of Artificial Intelligence ,'PHI Publication'.