Minimum Configuration MLP for Solving XOR Problem

Vaibhav Kant Singh
Assistant Professor, Department of Computer Science and Engineering, Institute of Technology, GGV,
Central University, Bilaspur, Chhattisgarh, India
Email Id:vibhu200427@gmail.com

Shweta Pandey
Master of Technology Student,
Indira Gandhi Delhi Technical University for Women,
Delhi, India
Email Id: shwetapandey9207@gmail.com

Abstract – There are various existing solution for solving the XOR problem in Artificial Neural Network. Even the author in his previous work has proposed several solutions to the problem. In this paper we will see a new solution to the XOR problem. The author has given a brief introduction to the Artificial Neural Network concept. Discussion on linear separablility and non-linear separability is made. Fixation of the non-linearly separable problem is made. A solution to the non-linearly separable problem i.e. XOR problem is proposed. Architectural Graph and Signal Flow Graph proposing the final solution to the problem is given. Mathematical explanation to the solution is given.

Keywords – ANN, AND, BNN, MLP, NAND, NOR, OR, XOR.

NOMENCLATURE

ANN(Artificial Neural Network), BNN(Biological Neural Network), MLP(Multilayer Perceptron).

I. INTRODUCTION

Aritificial Neural Network as the name implies considers the construction of man-made(not real) network of neurons. The term neurons is derived from the nerve cell present in the BNN. The nerve cells present in the BNN are responsible for making all the computations which Human body makes after attaining input from the environment. Human beings from the five basic senses present in the Human body takes input from the environment he is currently in. The five basic senses include sense of touch, sense of smell, sense of hearing, sense of sight (vision) and sense of taste. For recognition of inputs from these senses human body is having sensory organs. The sensory organs may be termed as receptors. The receptors take input and forward the input to the network of neurons. For processing of input received from the sense organs the human brain contain separate sections. Human brain is capable of performing parallel and distributed computing because of the presence of different sections in brain dedicated for different types of tasks. The working of neurons the computation elements present in the human brain follows the phenomena observed in elements present in Physics. In physics there is a theory that says that every object tries to be in stable state. For being in stable state the object that desires to be in stable state tries to give out the extra energy that it is having. A similar phenomenon occurs in BNN. The neurons in the neural network are at a resting potential before it attains a new input via the dendrites present in its structure. At resting potential the neuron is having membrane impermeable to the extra cellular which is positively charged. After attaining input if the value of potential bypasses the stability amount the neuron is said to fire and this changes the permeability which allows the outside +Na ions to get inside and cause chemical reaction which causes imbalance and the neuron tries to get back to the stable state. While trying to get back to the stable state the neuron modifies the synaptic strength present in the neuronal structure. This is basically responsible for the attaining of the knowledge acquired by the brain.

ANN deals with the construction of computer programs which are analogous to the working which we described in the above paragraph of this section.There are various advantages of using ANN technique. Artificial Neural Network considers historical data and learning algorithm to train the network thus previous knowledge is used for making interpretation about a novel data which is a nice approach of dealing with a new data. ANN because of its framework can deal with non-linear tasks like fraud detection etc. ANN systems are far more efficient in storage of data once get trained. The time required to develop an application on ANN framework is generally less when compared with the development on conventional mechanism. ANN uses sensitivity analysis to do the prediction which is a very good method for description of the input.

Some of the currently existing ANN systems include ART, LVQ, Cauchy Machine, Hamming Network, Hopfield Network, AM, BSB, CPN, CCN, Boltzmann Machine, ADALINE, MADALINE etc.

The three important component of any ANN model are Network Topology, Learning algorithm and Individual Neuron. All ANN systems are classified into three network architec-

978-9-3805-4421-2/16/$31.00 ©2016 IEEE

tures namely Single Layer Network, Multilayer Network and Recurrent Network. For representing the ANN on a piece of paper there are three basic representation techniques Block Diagram representation, Signal Flow Graph and Architectural Graph. In this paper we utilized Signal Flow Graph and Architectural Graph in the proposed work. The learning algorithms used to train the ANN can be broadly classified into two categories on the basis of presence of teaching element into Supervised and Unsupervised algorithm. Generally if the network is aware of the desired response for the given input the algorithm is Supervised and Unsupervised otherwise. The neuron present in the ANN may be classified into Linear and Nonlinear neuron on the basis of the nature exhibited by the activation function used in the ANN system.

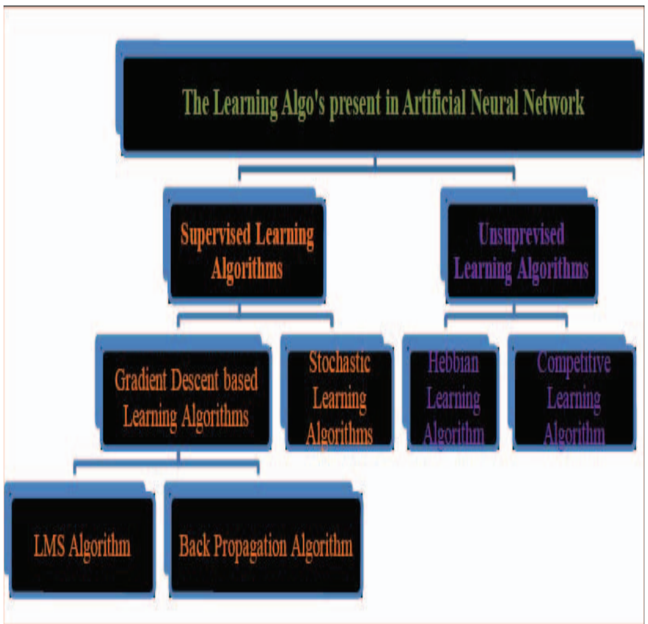

Fig. 1. Classification of the Learning Algorithms

II. LITERATURE SURVEY

In [1] Carpenter and Grossberg, gave a discussion on ART(Adaptive Resonance Theory) network. The ART employs a new principle of self organization. The complexity present in the ART networks is a limitation in its implementability. In [2,3] Rumelhart, Hinton and Williams gave Back propagation learning concept. In [4,5,6] Werbos introduced and discussed Backpropagation learning. In [7] Parker discussed second order backpropagation. In [2,3,4,5,6,7] Back propagation technique is utilized, thus the networks utilizing the algorithm are called Back propagation networks (BPN). Generally Backpropagation learning algorithms are used for Feed forward networks. Back propagation network suffers from the problem of local minima, which is a major problem in the networks employing the concept. Slow convergence rate is a problem in Backpropagation network. In [8] Kosko described BAM (Bidirectional Associative Memory). BAM are generally two layered recurrent, hetero associative networks. The network is capable of storing pattern pairs and on demand retrieving them. BAM behave like content addressable memory. In [9] Hinton and Sejnowski discussed Boltzmann machines. The Boltzmann machine are stochastic model whose state is governed by Boltzmann distribution. The Boltzmann machine suffers from the problem of computational load. In [10] Vaibhav Kant Singh gave the description of the problem of non-linear separabilty and provided one possible solution to the XOR problem wherein hyperbolic tangent function is used as activation function. In [11] Vaibhav Kant Singh proposed two more solutions to the XOR problem using hyperbolic tangent function as activation function. In [12] Vaibhav Kant Singh proposed a solution to the XOR problem where Logistic function is used as the activation function. In [17] Vaibhav Kant Singh proposed the solution to the construction of various logic GATES namely AND, OR, NOT, NAND and NOR. In [14] Vaibhav Kant Singh proposed solution to the Ex-NOR problem. In [18] Vaibhav Kant Singh gave brief introduction to some of the learning algorithms used for training ANN.

III. PROBLEM STATEMENT

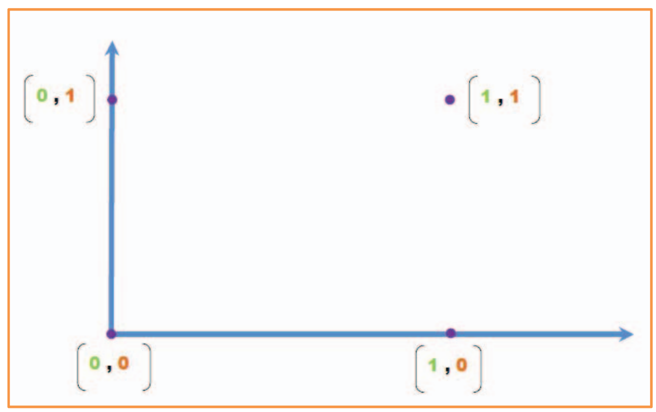

Fig. 2. Representation of Two inputs given to any two input Logic Gate (In 2 Dimension)

Figure 2 represents a two dimensional diagram wherein the set of possible inputs are represented along X and Y axis. The algorithms proposed for implementing the learning algorithms for solving problems can be classified into two categories on the basis of its feature of distinguishing or classifying the input into classes on the basis of output. The classes of algorithm may be termed as linearly separable and non-linearly separable problems. The author has given implementation of the various Logic Gates AND, NAND, NOR and OR which are linearly separable in his previous work. Figure 3 and Figure 4 shows that it is possible to draw a line between the Inputs {(0,0),(1,0),(0,1),(1,0)} applicable in the NAND, NOR, OR and AND GATES. There is a line in the Figure that classifies the inputs into two linearly separable classes. Class1 giving similar output and Class2 giving similar output. All the algorithms and models which follow this principle of operation come under Linearly Separable algorithms and Models. Beside to the above four GATES NOT GATE is also linearly separable.

The XOR problem discussed in the paper is a non-linearly separable problem as it is not possible to divide or classify the input into classes on the basis of output. The solution to the problem is termed as non-linearly separable solution. It was found in the previous study made by the author that it is not possible to solve the non-linearly separable problem using one layer.

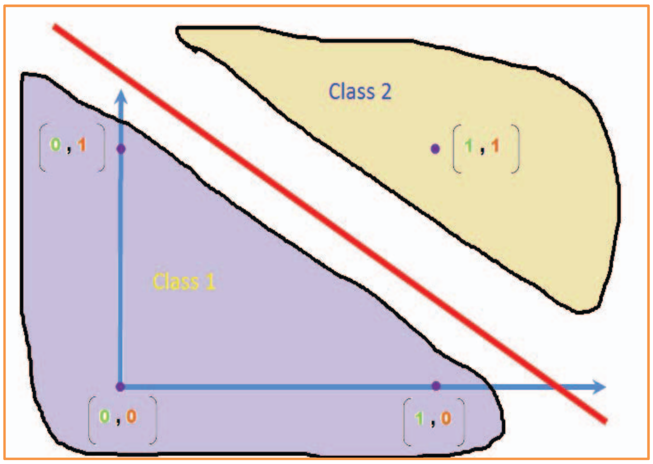

Fig. 3. Division of Inputs into two classes on the basis of Output for AND and NAND GATE

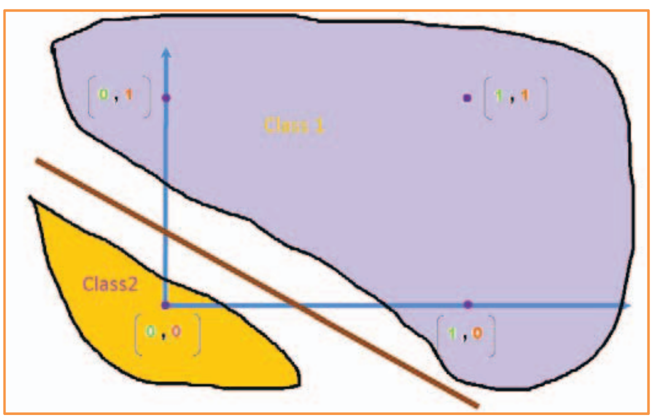

Fig. 4. Division of Inputs into two classes on the basis of Output for OR and NOR GATE

IV.EXISTING SOLUTIONS

XOR problem is being solved by various ANN systems.Vaibhav Kant Singh author of this paper has proposed various solutions to the problem [10,11,12] in his previous work in the year 2015. XOR problem solution was proposed by Touretzky and Pomerleau [14] in 1989. The proposed solution uses threshold function as activation function. The solution to the XOR problem can be obtained using the MADALINE ANN systems also. XOR problem can be solved using the Radial Basis Function Networks also given in Haykin [15]. XOR problem can be solved using Support Vector Machines taking basis from Cherkassky and Mulier [16] 1998. All the proposed solutions which we discussed in this section are having different operational architecture.

V.PROPOSED WORK

XOR problem which is a non-linearly separable problem could not be solved by the conventional methods of Single layer neural network systems. The solution proposed in this paper is through the minimum configuration MLP. The proposed solution contains the final values of the free parameters. Final values means the values at which the Artificial Neural Network provide desired values for the inputs. We know that in the minimum configuration MLP a number of hidden layers having non-linear element is possible with the output layer having linear element. In the solution there are two neurons in the hidden layer as represented by the Architectural Graph of the Fig. The Activation function used in the Hidden layer is Logistic function. The Activation function used in the output layer is Threshold function.

Before going in to the proposed solution it is important to note that the values of weight shown in the proposed model are the finally obtained values of weights obtained after performing the learning procedure to the ANN.

As we all know that MLP is a multilayer network thus is going to contain more than one layer. In the proposed solution there are two neurons in the hidden layer. Now, we will see the solution to XOR problem described in terms of Cases.

Case1
From Fig taking the value of x1=0 and x2=0 we get the value of "node1"as

$$\boldsymbol{node1 = x1 \times 1 + x2 \times -1 + 1 \times (-0.5) \ldots\ldots\ldots\ldots (1)}$$

Putting the value of x1 and x2 in Eq(1) we get

$$\boldsymbol{node1 = 0 \times 1 + 0 \times -1 + 1 \times (-0.5) = (-0.5) \ldots\ldots (2)}$$

From Fig the node value of "node2" will be

$$\boldsymbol{node2 = x1 \times -1 + x2 \times 1 + 1 \times (-0.5) \ldots\ldots\ldots\ldots (3)}$$

Putting the value of x1 and x2 in Eq(3) we get

$$\boldsymbol{node2 = 0 \times -1 + 0 \times 1 + 1 \times (-0.5) = (-0.5) \ldots\ldots (4)}$$

TABLE I. TABLE REPRESENTING XOR LOGIC

Logic for XOR problem		
x1	*x2*	$\boldsymbol{y = \overline{x1 \oplus x2}}$
0	0	0
1	0	1
0	1	1
1	1	0

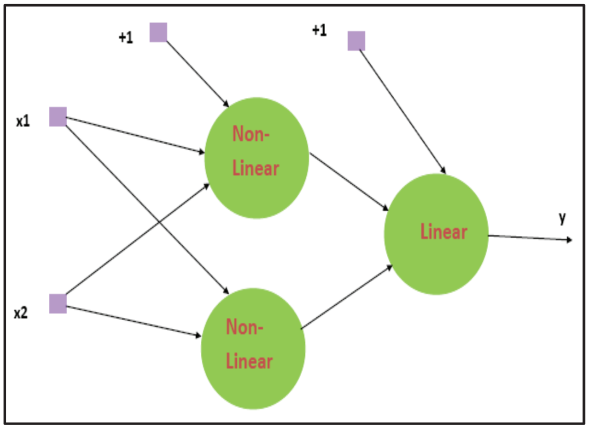

Fig. 5. Architectural Graph representing MLP Model for Solution

Here, the Activation function in the hidden layer is logistic function (non-linear elements in the hidden layer). The definition of logistic function Ă(x) is

$$Ă(\boldsymbol{x}) = \frac{\boldsymbol{L}}{\boldsymbol{1} + \boldsymbol{e}^{-\boldsymbol{k}(\boldsymbol{x}-\boldsymbol{x0})}} \dots\dots\dots\dots\dots\dots\dots (\boldsymbol{5})$$

Where, $\boldsymbol{e} = \boldsymbol{2.718281828456}$, i.e. The natural logarithm base (also known as Euler's number), x=Induced local field, $\boldsymbol{x0} = \boldsymbol{0}$ where x0 is the x-value of the sigmoids midpoint, $\boldsymbol{L} = \boldsymbol{1}$, where L is the Curve's maximum value, k is assumed to be 1 here k is steepness of the curve and Range[0,+1].

Therefore for this solution putting the values of x0, k and L in Eq(5) we get

$$Ă(\boldsymbol{x}) = \frac{\boldsymbol{1}}{\boldsymbol{1} + \boldsymbol{e}^{-\boldsymbol{x}}} \dots\dots\dots\dots\dots\dots\dots\dots\dots . (\boldsymbol{6})$$

From Fig the value of "node3" will be from Eq(6) and Eq(2) having the value of "node1"

$$\boldsymbol{node3} = Ă(\boldsymbol{node1}) = Ă(-\boldsymbol{0.5}) = \frac{\boldsymbol{1}}{\boldsymbol{1} + \boldsymbol{e}^{-(-\boldsymbol{0.5})}}$$

$$\boldsymbol{node3} = \frac{\boldsymbol{1}}{\boldsymbol{2.648721271}} = (\boldsymbol{0.377540668}) \dots (\boldsymbol{7})$$

From Fig the value of "node4" will be from Eq(6) and Eq(4) having the value of "node2"

$$\boldsymbol{node4} = Ă(\boldsymbol{node2}) = Ă(-\boldsymbol{0.5}) = \frac{\boldsymbol{1}}{\boldsymbol{1} + \boldsymbol{e}^{-(-\boldsymbol{0.5})}}$$

$$\boldsymbol{node4} = \frac{\boldsymbol{1}}{\boldsymbol{2.648721271}} = (\boldsymbol{0.377540668}) \dots\dots (\boldsymbol{8})$$

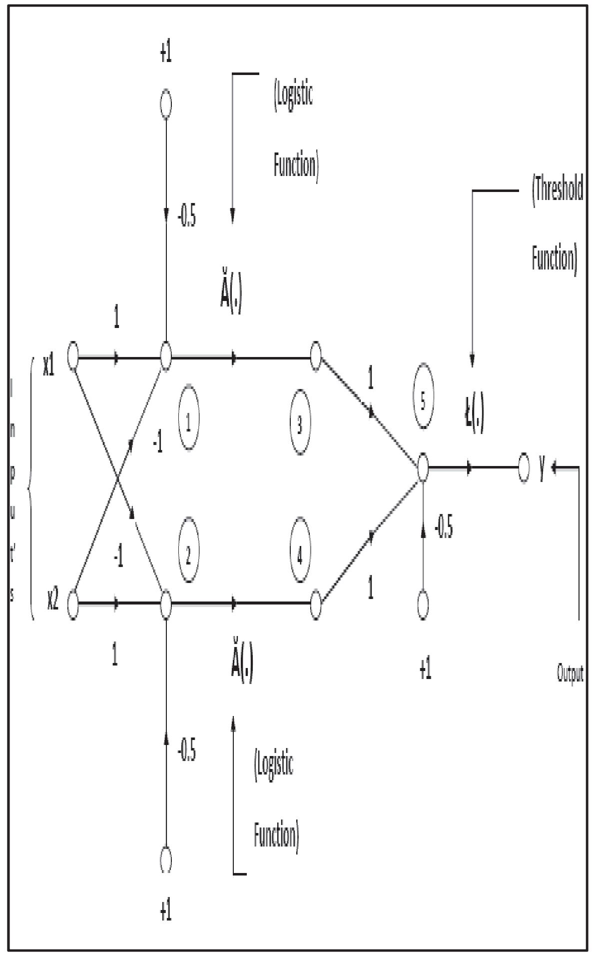

Fig. 6. Signal Flow Graph representing MLP Model for Solution

From Fig the "node5" value will be

$$\boldsymbol{node5} = \boldsymbol{node3} \times \boldsymbol{1} + \boldsymbol{node4} \times \boldsymbol{1} + \boldsymbol{1} \times (-\boldsymbol{0.5}) .. (\boldsymbol{9})$$

Putting the value of "node3" from Eq(7) and "node4" from Eq(8) in Eq(9) we get

$$\boldsymbol{node5} = (\boldsymbol{0.377540668} \times \boldsymbol{1}) + (\boldsymbol{0.377540668} \times \boldsymbol{1}) - \boldsymbol{0.5}$$

$$\boldsymbol{node5} = (\boldsymbol{0.255081336}) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots . (\boldsymbol{10})$$

Here, In this solution the Activation function in the output layer will be Threshold function. The definition is given as

$$Ł(\boldsymbol{x}) = \begin{cases} \boldsymbol{1\ if x} \geq \boldsymbol{threshold} \\ \boldsymbol{0\ if x} < \boldsymbol{threshold} \end{cases} \dots\dots\dots\dots\dots (11)$$

Here, the threshold value is "0.28". Therefore putting the value of threshold in Eq(11) we get

$$Ł(\boldsymbol{x}) = \begin{cases} \boldsymbol{1\ if x} \geq \boldsymbol{0.28} \\ \boldsymbol{0\ if x} < 0.28 \end{cases} \dots\dots\dots\dots\dots\dots\dots\dots\dots . (\boldsymbol{12})$$

Therefore, From Fig the value of y for x1=0 and x2=0 will be from Eq(12) and Eq(10) having the value of "node5"

$$y = Ł(node5) = Ł(0.255081336) = 0 \dots\dots (13)$$

$$Since, (0.255081336 < 0.28) where\ 0.28\ isthreshold$$

Case2

From Fig taking the value of x1=1 and x2=0 we get "node1" from Eq(1)

$$node1 = 1 \times 1 + 0 \times -1 - 0.5 = 0.5 \dots\dots (14).$$

From Fig substituting the value of x1 and x2 in Eq(3) we get "node2" as

$$node2 = 1 \times -1 + 0 \times 1 - 0.5 = -1.5 \dots\dots (15)$$

From Fig the value of "node3" will be from Eq(6) and value of "node1" in Eq(14) $node3 = Ă(node1) = Ă(0.5)$

$$node3 = \frac{1}{1+e^{-0.5}} = \frac{1}{1.60653066}$$

$$node3 = 0.622459331 \dots\dots (16)$$

From Fig the value of "node4" will be from Eq(6) and value of "node2" from Eq(15) $node4 = Ă(node2) = Ă(-1.5)$

$$node4 = \frac{1}{1+e^{1.5}} = \frac{1}{5.48168907}$$

$$node4 = 0.182425523 \dots . (17)$$

From Fig the value of "node5" will be from Eq(9), value of "node3" from Eq(16) and value of "node4" from Eq(17)

$$node5 = (0.622459331 \times 1) + (0.1824225523 \times 1) - 0.5$$

$$node5 = (0.30488485) \dots\dots\dots\dots\dots\dots\dots\dots (18)$$

From Fig the value of y for x1=1 and x2=0 will be from Eq(12) and Eq(18) having the value of "node5"

$$y = Ł(node5) = Ł(0.30488485) = 1 \dots\dots\dots\dots (19)$$

$$Since, (0.30488485 > 0.28) where\ 0.28\ isthreshold$$

Case3

From Fig taking the value of x1=0 and x2=1 we get "node1" from Eq(1) $node1 = 0 \times 1 + 1 \times -1 - 0.5 = -1.5 \dots (20)$.

From Fig taking the value of x1=0 and x2=1 we get "node2" from Eq(3) $node2 = 0 \times -1 + 1 \times 1 - 0.5 = 0.5 \dots\dots (21)$

From Fig the value of "node3" will be from Eq(6) and from the value of "node1" in Eq(20)

$$node3 = Ă(node1) = Ă(-1.5) = \frac{1}{1+e^{1.5}}$$

$$node3 = \frac{1}{5.48168907}$$

$$node3 = 0.182425523 \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (22)$$

From Fig the value of "node4" will be from Eq(6) and from the value of "node2" in Eq(21)

$$node4 = Ă(node2) = Ă(0.5) = \frac{1}{1+e^{-0.5}} = \frac{1}{1.60653066}$$

$$node4 = 0.622459331 \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (23)$$

From Fig the value of "node5" will be from Eq(9), value of "node3" in Eq(22) and value of "node4" in Eq(23)

$$node5 = (0.182425523 \times 1) + (0.62245933 \times 1) - 0.5$$

$$node5 = 0.30488485 \dots\dots\dots\dots\dots\dots\dots\dots\dots (24)$$

From Fig the value of y for x1=0 and x2=1 will be from Eq(12) and value of "node5" from Eq(24)

$$y = Ł(node5) = Ł(0.30488485) = 1 \dots\dots\dots\dots\dots (25)$$

$$Since\ 0.30488485 > 0.28\ where\ 0.28\ isthreshold$$

Case4

From Fig taking the value of x1=1 and x2=1 we get "node1" from Eq(1)

$$node1 = 1 \times 1 + 1 \times -1 - 0.5 = -0.5 \dots . (26)$$

From Fig taking the value of x1=1 and x2=1 we get "node2" from Eq(3)

$$node2 = 1 \times -1 + 1 \times 1 - 0.5 = -0.5 \dots . (27)$$

From Fig the value of "node3" will be from Eq(6) and value of "node1" in Eq(26) $node3 = Ă(node1) = Ă(-0.5)$

$$node3 = \frac{1}{1+e^{0.5}} = \frac{1}{2.648721271}$$

$$node3 = 0.377540668 \dots . (28)$$

From Fig the value of "node4" will be from Eq(6) and value of "node2" in Eq(27) $node4 = Ă(node2) = Ă(-0.5)$

$$node4 = \frac{1}{1+e^{0.5}} = \frac{1}{2.648721271}$$

$$node4 = 0.377540668 \dots . (29)$$

From Fig the value of "node5" will be from Eq(9), value of "node3" in Eq(28) and value of "node4" in Eq(29)

$$node5 = (0.377540668 \times 1) + (0.377540668 \times 1) - 0.5$$

$$node5 = 0.255081336 \dots\dots\dots\dots\dots\dots\dots\dots\dots (30)$$

From Fig the value of y for x1=1 and x2=1 will be from Eq(12) and value of "node5" Eq(30)

$$\boldsymbol{y} = \text{Ł}(\boldsymbol{node5}) = \text{Ł}(\boldsymbol{0.255081336}) = \boldsymbol{0} \ldots\ldots\ldots.(\boldsymbol{31})$$

$$Since\ 0.255081336 < 0.28\ where\ 0.28\ isthethreshold$$

VI. CONCLUSION AND FUTURE SCOPE

From the section proposed work it is clear that it is possible to construct a model of Minimum Configuration MLP to solve the problem of non-linear separablity present in the XOR problem. Multilayer Network can solve the problem of non-linear separability i.e. XOR problem is proved by this paper. The solution provides a platform for researchers working to find solution for non-linearly separable problems using the Minimum Configuration MLP model.

ACKNOWLEDGEMENT

First Author want to thank his father Shaheed Late Shri Triveni Kant Singh for his blessing that are always on him and which enabled him to do this work. First Author also want to place his sincere thanks to his mother Mrs. Sushma Singh, his wife Mrs. Shubhra Singh, his brothers Mr. Abhinav Kant Singh and Abhuday Kant Singh, his sister Miss Priyamvada Singh, his sweet daughter Miss Akshi Singh and his naughty son Master Aishwaryat Singh for their support and resources that they gave for the work that he did. First Author also want to thank his uncles Dr. Vinay Kumar Singh and Mr. Rishi Kumar Singh for the moral support that they always gave him. First author also want to thank his Grandmother for her blessings. Last but not the least First author also want to thank his Grandfather because of whom he is having existence.

REFERENCES

[1] G. A. Carpenter and S. Grossberg, "Adaptive resonance theory (ART)", in *The Handbook of Brain Theory and Neural Networks*, M. A. Arbib, Eds, Cambridge, MA:MIT press, 1995, pp. 79-82.

[2] D. E. Rumelhart, G. E. Hinton and R. J. Williams, "Learning representations of back-propagation errors", *Nature (London)*, Vol. 323, pp. 533-536, 1986.

[3] D. E. Rumelhart, G. E. Hinton and R. J. Williams, "Learning internal representations by error propagation", inD. E. Rumelhart and J. L. McCleland, Eds., Vol. 1, Chapter 8, Cambridge, MA:MIT press, 1986.

[4] P. J. Werbos, "Backpropagation through time:What it does and how to do it", In Proc. of IEEE, Vol. 78, 1990, pp. 1550-1560.

[5] P. J. Werbos, "Backpropagation and neurocontrol: A review and prospectus", In Proc. of International Joint Conference on Neural Networks, Washington, DC, Vol. I, 1989, pp. 209-216.

[6] P. J. Werbos, "Beyond regression:New tools for prediction and analysis in the behavioral sciences", *Ph.D. Thesis*, Harvard University, Cambridge, MA, 1974.

[7] D. B. Parker, "Optimal algorithms for adaptive networks: Second order backpropagation, second order direct propagation, and second order Hebbian learning", *IEEE 1st International Conference on Neural Networks*, Vol. 2, pp. 593 – 600, San Diego, CA, 1987.

[8] B. Kosko, "Bidirectional Associative Memories", *IEEE Transactions onSystems, Man, and Cybernetics*, Vol. 18, pp. 49 – 60, 1988.

[9] G. E. Hinton, and T. J. Sejnowski, "Learning and relearning in Boltzmann Machines", in *Parallel and Distributed Processing: Explorations in Microstructures of Cognition*, D. E. Rumelhart and J. L. McClelland, eds., Cambridge, MA:MIT Press, 1986.

[10] V.K. Singh, "One Solution to XOR problem using Multilayer Perceptron having Minimum Configuration", *International Journal of Science and Engineering*, Vol. 3, No. 2, pp. 32 – 41, 2015.

[11] V.K. Singh, "Two Solutions to the XOR problem using minimum configuration MLP", *International Journal of Advanced Engineering Science and Technological Research*, Vol. 3, No. 3, pp. 16 – 20, September 2015.

[12] V.K. Singh, "Proposing Solution to XOR problem using minimum configuration MLP", Accepted for Publishing in Procedia Computer Science, *International Conference on Computational Modeling and Security (CMS 2016),Publisher Science Direct Elsevier,* R.L. Jalappa Institute of Technology, Bangalore, India, February 11-13 2016.

[13] V.K. Singh, "Mathematical Explanation to solution for EX-NOR problem using MLFFN", *Published in National Conference NCSCA-2016,* Smt. ChandabenMohanbhai Patel Institute of Computer Appilications, Charotar University of Science and Technology, Changa, Anand, Gujarat, India, February 5-6, 2016.

[14] D.S. Touretzky, and D.A. Pomerleau, "What is hidden in the hidden layers?" *Byte*, Vol 14, pp. 227-233, 1989.

[15] S. Haykin, *Neural Networks A Comprehensive Foundation*, Low Price Edition, Pearson Education Asia, Singapore, 2002.

[16] V. Cherkassky and F. Mulier, *Learning from Data : Concepts*, Theory and Methods, Wiley:Newyork, 1998.

[17] V.K. Singh, "ANN Implementation of Constructing Logic Gates Focusing On Ex-NOR," *Published in Proceeding of BITCON National Conference*, BIT, Durg, India, January 29-30, 2016.

[18] V.K. Singh, "Mathematical Analysis For Training ANNs Using Basic Learning Algorithms,"*Published in Proceeding of BITCON National Conference*, BIT, Durg, Chhattisgarh, India, January 29-30, 2016.