Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 85 (2016) 263 – 270

Procedia Computer Science

International Conference on Computational Modeling and Security (CMS 2016)

Proposing Solution to XOR problem using minimum configuration MLP

Vaibhav Kant Singh*

Assistant Professor, Department of Computer Science & Engineering, Institute ot Technology, Guru Ghasidas Vishwavidyalaya, Central University, Bilaspur, Chhattisgarh, Pincode 495001, India

Abstract

This paper is an extension to what the author had already done in [1] and [2]. In this paper we will see one solution to the XOR problem using minimum configuration MLP an ANN model. The proposed solution is proved mathematically in this paper. The problem of non-linear separability is addressed in the paper. The Architectural Graph representation of the proposed model is placed and also an equivalent Signal Flow Graph is represented to show how the proof the proposed solution. The non-linear Activation function used for the hidden layer minimum configuration MLP is Logistic function.

*Keywords:*ANN, XOR, MLP;

1. Introduction

McCulloch who was a psychiatrist after making research in the behavior of human brain towards various stimuli collaborated with Pitts who was a mathematician. In the year 1943 they proposed [11] a model in their paper using the logical calculus representing neural network. The model united the studies of neurophysiology and mathematical logic. The paper was very widely studied during that time and influenced von Neumann to derive EDVAC. Since,

* Corresponding author. Tel.: +919424174443; fax: +0-000-000-0000 .
*E-mail address:*vibhu200427@gmail.com

doi:10.1016/j.procs.2016.05.231

then various neural network models were proposed. In this paper we will see a problem which was encountered while making solution for various problem domain using the models.

1.1. Structure and working of Biological neuron

Human brain is composed of cells which are called nerve cells or neurons. The basic structure of neuron comprises of dendrites responsible for taking inputs, soma (the cell body), mylelin shealth, axon the transmission link and Synapses responsible for storing the acquired knowledge is shown in Fig.1. In the 18th Century Galvani showed that there is a potential difference between across the membrane. Before acquiring any input from outside the cell is at a resting potential. If after attaining input through dendrite the potential changes and neuron fires certain neurotransmitters are produced as the cell membrane becomes impermeable and creates a scenario where chemical reaction takes place. These neurotransmitters produced are responsible for making alteration in the Synaptic junction and thus responsible for recognizing an stimuli.

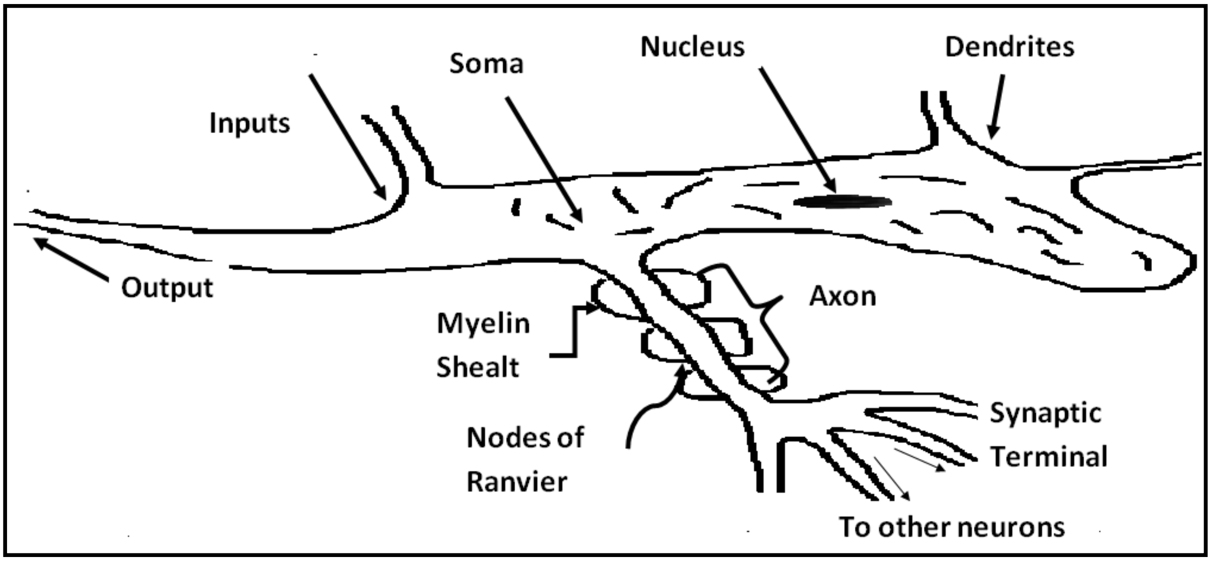

Fig. 1. Structure of Biological Neuron

1.2. Artificial Neural Network.

Artificial Neural Network is a parallel and distributed processor that is modeled to perform task in a similar manner to the working of human brain. Looking to the capability of human brain that solves wide variety of problem ANN was modeled. ANN is having the capability of computer system to solve the problem. ANN is comparatively faster than human brain processing. ANN is going to perform the task several times without getting annoyed. There are three fundamental elements of an ANN namely the individual artificial neuron, learning rule and network topology that connect the neurons. There are three basic techniques utilized for representing ANN on a piece of paper. The representation techniques are Block diagram representation, signal flow graph and architectural graph. Learning algorithms are basically of three types Supervised, Unsupervised and Reinforced learning. There are three basic network architectures Single layer Feedforward, Multilayer Feedforward and Recurrent Network. Neuron used may be linear or non linear depending upon the type of activation function used inside it. In this paper example of all three is present Fig 2 represents Block diagram, Fig 3, Fig 4 and Fig 5 represent Architectural Graph whereas Fig. 6 represents Signal flow graph proposed by Mason in 1953. Some of the neural network models are ADALINE, MADALINE, ART, AM, BAM, CCN, Boltzmann machine, BSB, Hopfield network, Hamming network, LVQ, CPN, RNN etc. In this paper we will see two models i.e. Perceptron Model and Multilayer Perceptron Model.

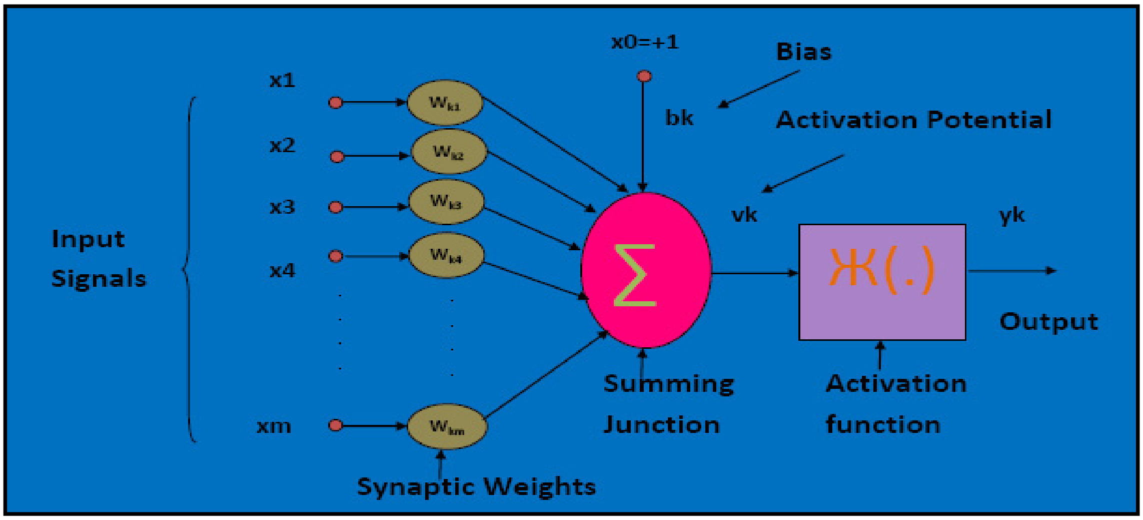

Fig. 2. Block Diagram representation of a neuron in ANN

2. Perceptron Model

In 1958 Frank Rosenblatt [12] proposed the Perceptron Model which he named as brain model. Perceptron model was used for solving pattern classification problems. Perceptron was the first model which was making use of supervised learning for training the network. The model mostly comprise of single neuron with adjustable weights and bias. Rosenblatt in his study proved that if the vectors used to train the network are taken from linearly separable classes then the algorithm which he named perceptron convergence algorithm will converge and position the decision surface in the form of hyper plane between the two classes. The proof of convergence is known as Perceptron Convergence Theorem. The Activation function used in the neuron is signum (sgn). The definition is:-

$$sgn(v) = \begin{cases} +1\ if\ v > 0 \\ -1\ if\ v < 0 \end{cases}\ here, v\ is\ the\ induced\ local\ field\ value$$

The Equations for Update weight vector w(n+1), desired response d(n) and the actual response y(n)is given below:-

$$y(n) = sgn\big(w\text{Ŧ}(n)x(n)\big), where\ y(n) is\ actual\ response, x(n) is\ the\ input\ vector\ and\ w\text{Ŧ}(n) is\ weight\ matrix$$

$$w(n+1) = w(n) + \sigma[d(n) - y(n)]x(n), where\ \sigma\ is\ learning\ constant\ and\ n\ denotes\ the\ iteration\ step$$

$$d(n) = \begin{cases} +1\ if\ x(n) belons\ to\ Class\ €1 \\ -1\ if\ x(n) belons\ to\ Class\ €2 \end{cases}\ where\ d(n) represents\ desired\ response$$

The perceptron model could solve only the problems which are linearly separable. Since there are various problems which are inherently non-linearly separable the solution was not possible through the perceptron model. The perceptron model was of single layer although there could be more than one neuron in the layer.

3. Problem Statement

Two sets of points A and B in the n-dimensional space are called linearly separable if (n+1) real numbers $w_1, \ldots\ldots w_{n+1}$ exist, such that every point $(x1, x2, \ldots \ldots . xn) \in A$ satisfies $\sum_{i=1}^{n} wixi \geq w_{n+1}$ exist, and every point $(x1, x2 \ldots \ldots \ldots xn) \in B$ satisfies $\sum_{i=1}^{n} wixi < w_{n+1}$. The problems which satisfy the above statement are called linearly separable problem. The construction of logic gates AND, OR, NAND, NOR, NOT using single layer models is possible because of the presence of linear separabilty in the problems. The problems that do not satisfy the above statement are called non linearly separable problem. The XOR problem discussed in this paper is a non linearly separable problem. It is not possible to solve the XOR problem using the single layer model because of presence of non linearity in the problem exhibited by XOR logic.The discussion of non linear separabilty exhibited by XOR is discussed by the author in [1].

4. Literature Survey

In [1] and [2] the author has proposed four different solutions to the XOR problem using the minimum configuration MLP. The Activation function used in those solutions was hyperbolic tangent function. In 1989 Touretzky and Pomerleau [3] proposed one solution to XOR problem using threshold function as Activation function. In [3] the one solution to the XOR problem is given using RBF network which supports only one hidden layer. In 1998 Cherkassky and Mulier [4] proposed solution to the XOR problem using Support Vector Machine concept. Polynomial machine is used to solve the problem. In [5] Anderson specified simple neural models using cognitive computation approaches are presented that will associate and will respond to prototypes of sets of related inputs. In [6] Abu-Mostafa and Jacques found that the asymptotic information capacity of a Hopfield network of N neurons is of the order N^3b. The number of arbitrary state vectors that can be made stable in a Hopfield network of N neurons is proved to be bounded above N. In [7] Amari made an attempt is made to establish a mathematical theory that shows the intrinsic mechanism, capabilities, and limitations of information processing by various architectures of neural networks. A method of statistically analyzing one-layer neural networks is given, covering the stability of associative mapping and mapping by totally random networks. In [8] Amari paid special attention on nets of randomly connected excitatory and inhibitory elements. In [9] Amari proposed error-correction adjustment procedures for determining the weight vector of linear pattern classifiers under general pattern distribution. In [10] Atiya and Abu-Mostafa proposed a method for the storage of analog vectors. He proved that the two layer case is guaranteed to store any number of given analog vectors provided their numbers does not exceed 1+ the number of neurons in the hidden layer.

5. Multilayer Perceptron

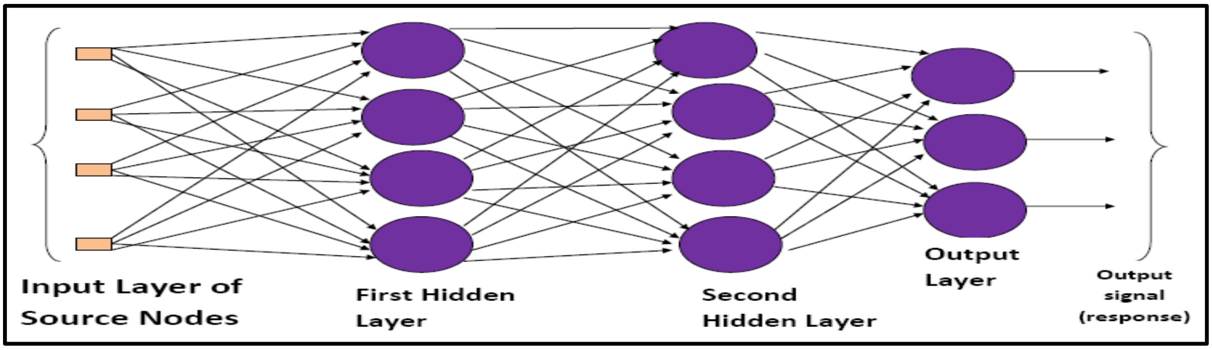

Fig. 3 Architectural Graph representing Multilayer Perceptron

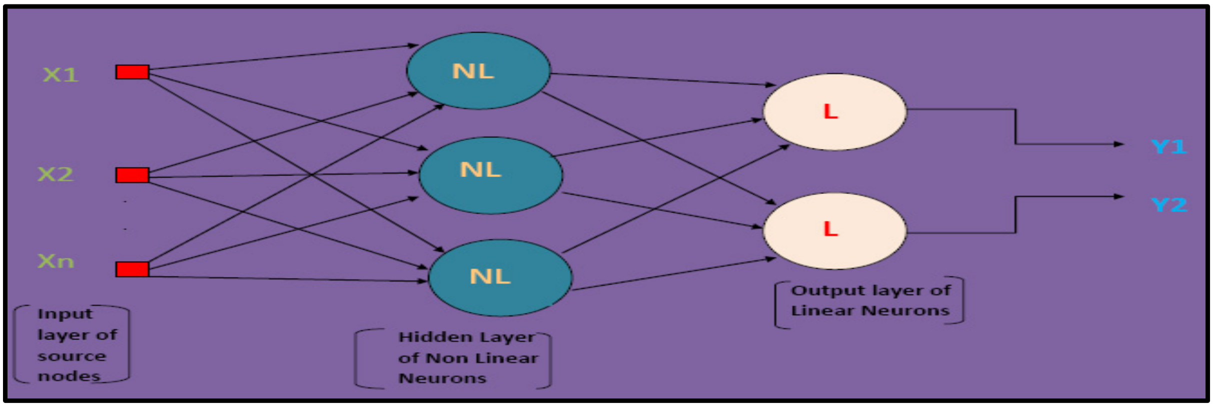
X1
X2
Xn
NL
NL
NL
L
L
Y1
Y2
Input layer of source nodes
Hidden Layer of Non Linear Neurons
Output layer of Linear Neurons

Fig. 4 Architectural Graph representing minimum configuration Multilayer Perceptron

Fig 3 represents MLP. MLP exhibits three basic characteristics which are (a) The model of neuron in the network includes non linear activation function. However a special case is the minimum configuration MLP, (b) The hidden layers present in the network have inherent capacity to learn complex patterns present in the input pattern and (c) The network connectivity present in the MLP is of high degree. MLP uses supervised learning algorithm called Backpropagation to train the network.

6. Proposed Solution

In this paper we will see the use of minimum configuration MLP shown in Fig 4. for solving the XOR problem which is a non-linearly separable problem. In minimum configuration MLP there could be any number of hidden layers comprising of non linear elements whereas in the output layer only linear neurons are present. All the neurons present in the hidden layer are non linear and all the neurons present in the output layer are linear. In the proposed solution where we have specified values for the weights present in the network. The specified values will enable to the minimum configuration MLP to give the desired output for all the possible inputs of XOR.

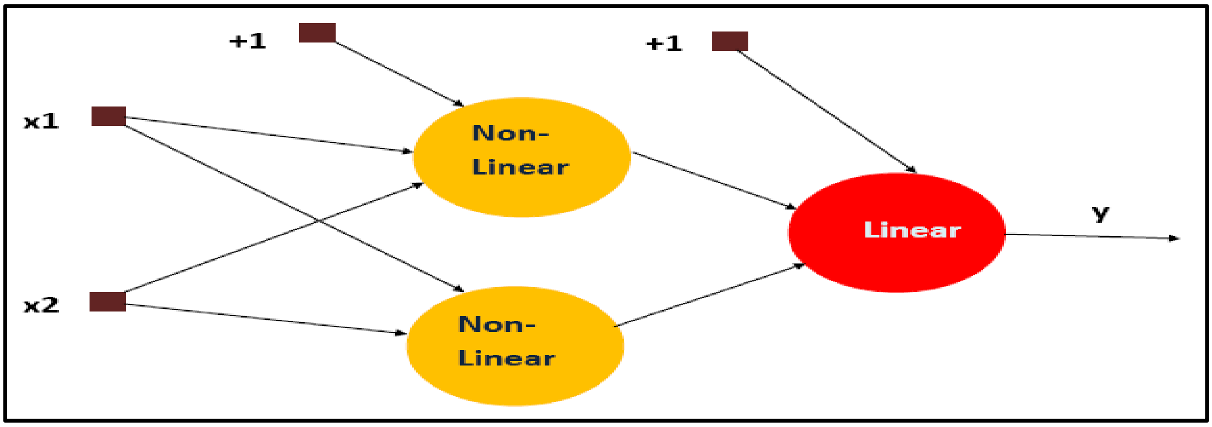

Fig. 5 Architectural Graph Representing the Solution to the XOR problem using minimum configuration MLP.

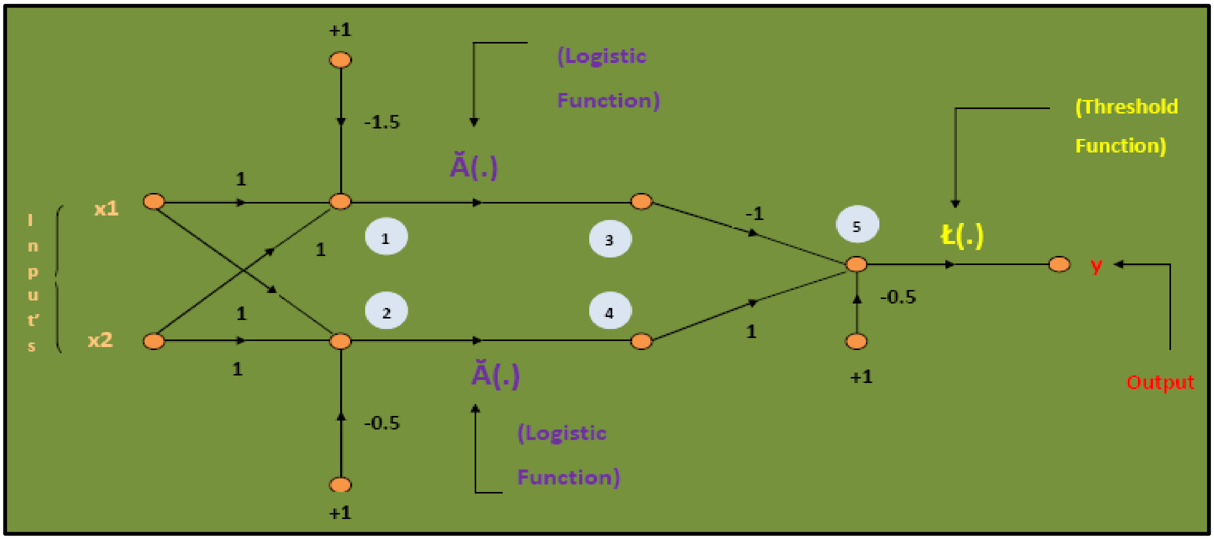

Fig.6 Signal Flow Graph Representing the Solution to the XOR problem using minimum configuration MLP.

From Fig.5 which represents architectural graph of the MLP provided as a solution to the XOR problem. In the proposed solution in the hidden layer there are two neurons. The neurons in the hidden layer are using logistic function as Activation function, whereas in the output layer there is one neuron. The output neuron is having Threshold function as Activation function. Fig.6 represents Signal flow graph representing the internal configuration of the neurons present in the Architectural Graph. In this section we will see mathematical solution to the XOR problem using minimum configuration MLP an ANN model. The mathematical solution will have 4 Cases as in the training set there will be four training pair. The training pair will be having $[x1 \quad x2]$as input vector and $[y]$ is the target vector.

Table 1.XOR Logic

$y = \overline{x1 \oplus x2}$	x1	x2
0	0	0
1	1	0
1	0	1
0	1	1

6.1. Case1

From Fig6 taking value of x1=0 and x2=0 the value of “node1” will be $node1 = x1 \times 1 + x2 \times 1 + 1 \times -1.5 \ldots\ldots\ldots (A)$

Substituting the value of x1 and x2 in Eq(A) we get $node1 = 0 \times 1 + 0 \times 1 - 1.5 = -1.5 \ldots\ldots\ldots (B)$

From Fig6 the value of “node2” will be $node2 = x1 \times 1 + x2 \times 1 + 1 \times -0.5 \ldots\ldots\ldots. (C)$

Substituting the value of x1 and x2 in Eq(C) we get $node2 = 0 \times 1 + 0 \times 1 - 0.5 = -0.5 \ldots\ldots\ldots\ldots (D)$

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

$$\text{Ă}(x) = \frac{L}{1 + e^{-k(x - x0)}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots (E)$$

Where, $e = 2.718281828456$, i.e. The natural logarithm base (also known as Euler’s number), x=Induced local field, $x0 = 0$ where x0 is the x-value of the sigmoids midpoint, $L = 1$, 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

$$\text{Ă}(x) = \frac{1}{1 + e^{-x}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots. (F)$$

From Fig6 the value of “node3” will be from Eq(F) and value of “node1” in Eq(B)

$$node3 = \text{Ă}(node1) = \text{Ă}(-1.5) = \frac{1}{1 + e^{1.5}} = \frac{1}{5.48168907} = 0.182425523 \ldots\ldots (G)$$

From Fig6 the value of “node4” will be from Eq(F) and value of “node2” in Eq(D)

$$node4 = \text{Ă}(node2) = \text{Ă}(-0.5) = \frac{1}{1 + e^{0.5}} = \frac{1}{2.648721271} = 0.377540668 \ldots. (H)$$

From Fig6 the value of “node5” will be $node5 = node3 \times -1 + node4 \times 1 + 1 \times (-0.5) \ldots (I)$

Putting the values of “node3” from Eq(G) and “node4” from Eq(H) in Eq(I) we get the value of “node5” as

$$node5 = (0.182425523 \times -1) + (0.377540668 \times 1) - 0.5 = -0.304884855 \ldots\ldots\ldots\ldots\ldots\ldots\ldots. (J)$$

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

$$\text{Ł}(x) = \begin{cases} 1 \ if \ x \geq threshold \\ 0 \ if \ x < threshold \end{cases} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. (K)$$

Here, the threshold value is “-0.28”. Therefore putting the value of threshold in Eq(K) we get

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

Therefore from Fig6 the value of y for x1=0 and x2=0 will be from Eq(L) and value of “node5” from Eq(J)

$$y = Ł(node5) = Ł(-0.304884855) = 0 \dots\dots\dots.(M)$$

$$Since - 0.304884855 < -0.28, here\ -0.28\ is\ threshold$$

6.2. Case2

From Fig6 taking value of x1=1 and x2=0 and substituting it in Eq(A) to get the value of “node1”.

$$node1 = 1 \times 1 + 0 \times 1 + 1 \times -1.5 = -0.5 \dots\dots..(N)$$

From Fig6 substituting the value of x1 and x2 in Eq(C) to get “node2” value as $node2 = 1 \times 1 + 0 \times 1 - 0.5 = 0.5 \dots(O)$

From Fig6 the value of “node3” will be from Eq(F) and value of “node1” in Eq(N)

$$node3 = Ă(node1) = Ă(-0.5) = \frac{1}{1 + e^{0.5}} = \frac{1}{2.648721271} = 0.377540668 \dots.(P)$$

From Fig6 the value of “node4” will be from Eq(F) and value of “node2” in Eq(O)

$$node4 = Ă(node2) = Ă(0.5) = \frac{1}{1 + e^{-0.5}} = \frac{1}{1.60653066} = 0.622459331 \dots\dots(Q)$$

From Fig6 the value of “node5” will be from Eq(I), value of “node3” in Eq(P) and value of “node4” in Eq(Q)

$$node5 = (0.377540668 \times -1) + (0.622459331 \times 1) - 0.5 = node5 = -0.255081337 \dots\dots\dots\dots\dots\dots\dots\dots\dots.(R)$$

From Fig6 the value of y for x1=1 and x2=0 will be from Eq(L) and value of “node5” from Eq(R)

$$y = Ł(node5) = Ł(-0.255081337) = 1 \dots\dots\dots\dots\dots(S)$$

$$Since - 0.255081337 > -0.28, here - 0.28\ is\ threshold$$

6.3. Case3

From Fig6 taking the value of x1=0 and x2=1 and substituting it in Eq(A) we get the value of “node1” as

$$node1 = 0 \times 1 + 1 \times 1 - 1.5 = -0.5 \dots\dots..(T)$$

From Fig6 taking value of x1=0 and x2=1 and substituting the value in Eq(C) we get the “node2” value as

$$node2 = 0 \times 1 + 1 \times 1 - 0.5 = 0.5 \dots\dots\dots\dots\dots.(U)$$

From Fig6 the value of “node3” will be from Eq(F) and value of “node1” in Eq(T)

$$node3 = Ă(node1) = Ă(-0.5) = \frac{1}{1 + e^{0.5}} = \frac{1}{2.648721271} = 0.377540668 \dots(V)$$

From Fig6 the value of “node4” will be from Eq(F) and value of “node2” in Eq(U)

$$node4 == Ă(node2) = Ă(0.5) = \frac{1}{1 + e^{-0.5}} = \frac{1}{1.60653066} = 0.622459331 \dots(W)$$

From Fig6 the value of “node5” will be from Eq(I), “node3” value from Eq(V) and “node4” value from Eq(W)

$$node5 = (0.377540668 \times -1) + (0.622459331 \times 1) - 0.5 = -0.255081337 \dots\dots\dots\dots\dots\dots\dots\dots(X)$$

From Fig6 the value of y for x1=0 and x2=1 will be from Eq(L) and value of “node5” from Eq(X)

$$y = Ł(node5) = Ł(-0.255081337) = 1 \dots\dots\dots\dots(Y)$$

$$Since - 0.255081337 > -0.28, here - 0.28\ is\ threshold$$

6.4. Case 4

Fig6 taking the value of x1=1 and x2=1 and substituting the value in Eq(A) to have the value of “node1” as

$$node1 = (1 \times 1 + 1 \times 1 - 1.5 = 0.5 \ldots\ldots\ldots\ldots\ldots (AA)$$

From Fig6 taking the value of x1=1 and x2=1 and substituting the value in Eq(C) we get the value of “node2” as

$$node2 = 1 \times 1 + 1 \times 1 - 0.5 = 1.5 \ldots\ldots\ldots\ldots\ldots\ldots . (BB)$$

From Fig6 the value of “node3” will be from Eq(F) and value of “node1” in Eq(AA)

$$node3 = \breve{A}(node1) = \breve{A}(0.5) = \frac{1}{1 + e^{-0.5}} = \frac{1}{1.60653066} = 0.622459331 \ldots (CC)$$

From Fig6 the value of “node4” will be from Eq(F) and value of “node2” in Eq(BB)

$$node4 = \breve{A}(node2) = \breve{A}(1.5) = \frac{1}{1 + e^{-1.5}} = \frac{1}{1.22313016} = 0.817574476 \ldots (DD)$$

From Fig6 the “node5” value will be from Eq(I), value of “node3” from Eq(CC) and value of “node4” from Eq(DD)

$$node5 = (0.622459331 \times -1) + (0.817574476 \times 1) - 0.5 == -0.304884855 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots . (EE)$$

From Fig6 the value of y for x1=1 and x2=1 will be from Eq(L) and value of “node5” obtained in Eq(EE)

$$y = Ł(node5) = Ł(-0.304884855) = 0 \ldots\ldots\ldots .. (FF)$$

$$Since - 0.304884855 < -0.28, here - 0.28\ is\ threshold$$

7. Conclusion

From Eq(M), Eq(S),Eq(Y) and Eq(FF) derived from the four cases in the above section it is proved that the proposed model solves the XOR problem using minimum configuration MLP. In this solution the variation that the author has made from his previous work is the use of Logistic function as Activation function for the hidden layers. Also from the work that we did it is proved that the problems which are non linearly separable could be solved using Multilayer Network. Minimum configuration MLP is a very nice approach for solving non linearly separable problem on account of its flexibility of using both linear and non linear neurons in the network.

References

1. Singh V.K., One solution to XOR problem using multilayer perceptron having minimum Configuration, To appear in International Journal of Science and Engineering 2015;3(2).
2. Singh V.K., Two solutions to the XOR problem using minimum configuration MLP, International Journal of Advanced Engineering Science and Technological Research 2015;3(3) p. 16-20.
3. Haykin S. *Neural networks a comprehensive foundation*, 2nd ed. Singapore: Pearson Education; 2002.
4. Cherkassky V., Mulier F., Self-organization as an iterative kernel smoothing process, Neural Computation, 1995;7 p. 1165-1177.
5. Anderson J.A., Cognitive and psychological computation with neural models, IEEE Transactions on Systems, Man and Cybernetics, 1983;13: p.799-815.
6. Abu-Mostafa Y.S., St. Jacques J.M., Information capacity of the hopfield model, IEEE Transactions on Information Theory, 1985; 31: p. 461-464.
7. Amari S., Mathematical foundations of neurocomputing, Proceeding of the IEEE, 1990; 78: p. 1443-1463.
8. Amari S., Characteristics of random nets of analog neuron like elements, IEEE Transactions on Systems, Man and Cybernetics, 1972; 2: p. 643-657.
9. Amari S., A theory of adaptive pattern classifiers, IEEE Transactions Electronic Computers, 1967; 16:p. 299-307.
10.Atiya A.F., Abu-Mostafa Y.S., An analog feedback associative memory, IEEE Transactions on Neural Networks, 1993; 4:p. 117-126.
11.McCulloch W.S.,Pitts W., A logical calculus of the ideas immanent in nervous activity, Bulletin of Mathematical Biophysics, 1943; 5:p. 115-133.
12. Rosenblatt F., The perceptron:A probabilistic model for information storage and organization in brain, Psychological Review, 1958; 65:p. 386-408.