IJSETT
International Journal

"International Journal for Science and Emerging Technologies with Latest Trends" 2(1): 20-28 (2012)

Role of Neural network in data mining

Chitranjanjit kaur
Associate Prof
Guru Nanak College,
Sukhchainana
Phagwara,(GNDU)
Punjab, India

Pooja kapoor
Associate Prof
Swami Sarvanand Group Of
Institutes
Dinanagar(PTU)
Punjab, India

Meenu Bala
Associate Prof
MMU, Mullana
Haryana, India

Abstract

Data Mining means mine data from huge amount of data. It is beneficial in every field like business, engineering, web data etc. In data mining classification of data is very difficult task that can be solving by using different algorithms. In this paper different neural networks are explain that help in classification of data and help in made supervised and unsupervised learning. This paper appreciates application of neural network in area of data mining.

Keywords: Data Mining, Single layer perceptron, Supervised, Unsupervised, Multilayer perceptron, Back propagation, Feed forward, Recurrent.

1.Introduction

Data mining is also known as Knowledge Discovery Data (KDD). It analyze large amount of data. It has relationship with other areas like neural network, database, and business intelligence. There are different type of learning mechanisms in the data mining supervised and unsupervised learning. Supervised learning means classification of data and unsupervised learning means clustering. Different methods are used to classify the data in data mining like decision trees, nearest neighbor, neural network. Neural network play significant role in data mining. Neural network consist of different node with weighted inputs, it is constructive in classification of complex data. Advantage of data mining is that it can construct and learn boundaries for large number of attributes. In this paper different algorithms are explain with examples that help in construct classification in data mining.

2.Learning system by Neural Networks:

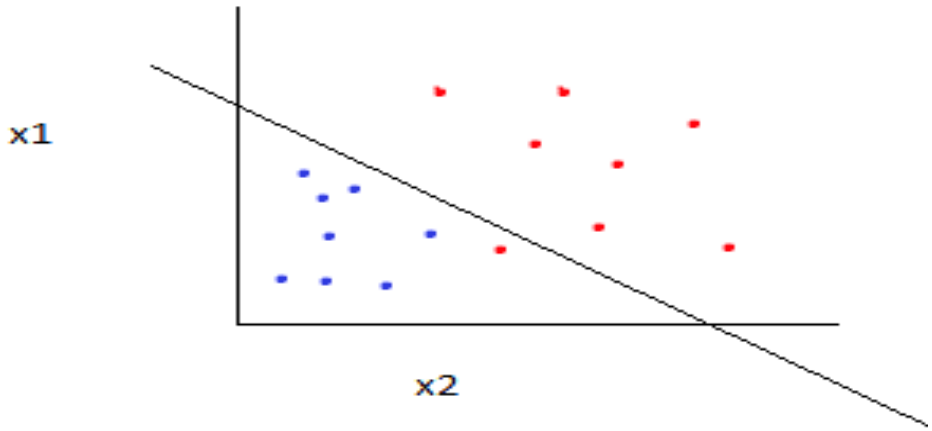
In data mining two learning methods are used one is classification and other is clustering. Classification mean Supervised learning in which data can determine by predetermine method. This learning system is predictive. There are different techniques for this naïve Bayes, decision tree and neural networks**.** For example on the base of old data we can predict eligibility of person for particular job. With neural network we can predict the value and find error between actual and desire outputs then adjust the new weight. In supervised learning different methods of neural network are there they are backpropagation, hopified, and recurrent method. Unsupervised

learning is known as clustering, it is not depend on the historic data but this method is depend on the good examples from the similar data. In this grouping of data is done. Different types of clustering methods are there partitional clustering on base of distance by k-mean, by density and hierarchal.

3 Classification by different types of Neural Networks

3.1 Single layer Neural Network:

It is also known as single layer perceptron. It is design by Frank Rosenblat. It is sum of all weighted inputs and doesn't require any prior knowledge. It is used to classify two classes by linear separable method and on the base of zero or one. In data mining it is use to classify two kind of data by decision boundary in two classes. Its example is below:

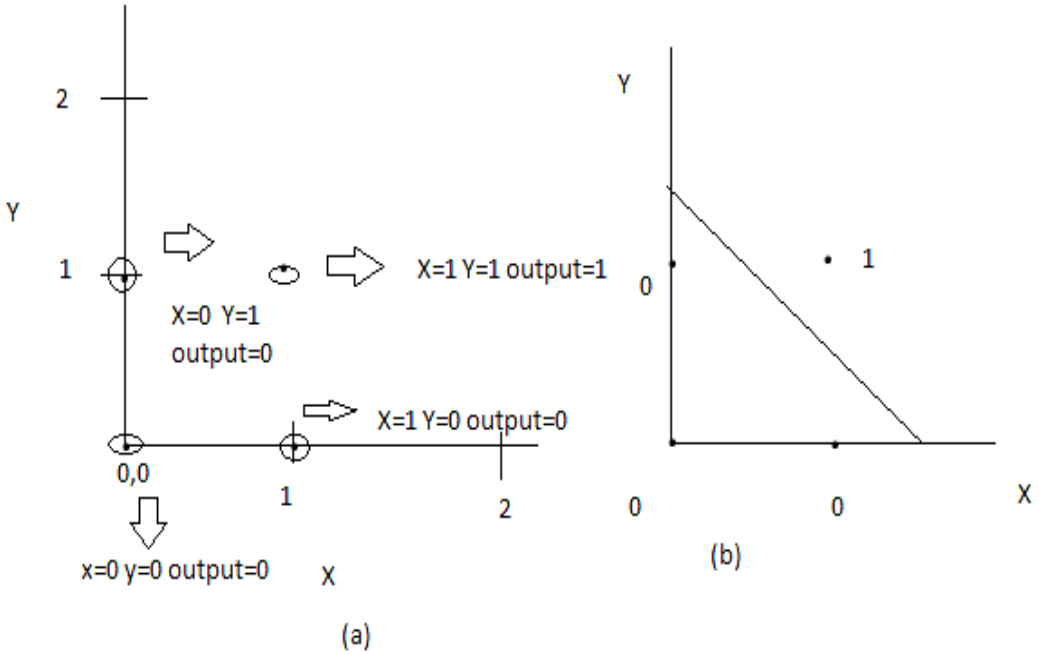


Red dots denote one class and blue dots refer to other class. Single perceptron classify two types of patterns. Single layer perceptron in data mining work for OR, AND and COMPLEMENT functions. It represent as follows. Example of AND function:

X	Y	Output
0	0	0
0	1	0
1	0	0
1	1	1

It represent as follows: In fig (a) representation of AND function is there when both inputs are true then output is true otherwise it is false. In fig (b) two classes are classify one with output 1 and other with output zero the class the line that divide two classes is decision boundary. This can be work in data mining for making decision between two classes as example below. Neural network is represent as below:

Basic steps in neural network are, first of all



som e

basic terms used in this

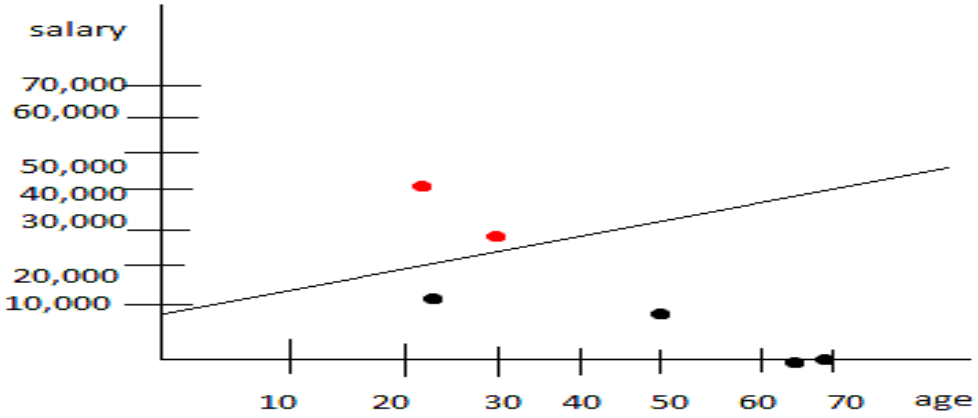
1. Inputs: In diagram shown below, the inputs are x0,x1
2. Bias: Bias is function that give input as one.
3. Summation: It is represent as ∑ symbol i.e summation of weighted inputs.
4. Activation function: this is function that decide weither output will be one or zero.
5. Desion boundary: It will classify two classes, as in below diagram decision boundary will be (x0*w0+x1*w1+w2=0). And two classes are (one is x0*w0+x1*w1+x2>0 and second x0*w0+x1*w1+w2<=0).

Some basic steps are:

1 Value of Input may be multidimension I=(x0,x2……..xn)
2 Input value of weights W=(w0,w1,w2……..wn)
3 Threshold=t
4 Summation weighted inputs may be taken p=∑ wi*xi
5 If p>0 , Then output=1
6 Else output=0
7 But if classification is incorrect then
wi= wi+ class[i]*xi
and then calculate again with new wi

S.no	Name	Age	Salary	Bank loan
1	Ram	50	10,000	No
2	Purnima	30	30,000	Yes
3	Menu	23	15,000	No
4	Jeeti	69	0	No
5	Pooja	74	0	No
6	Purab	25	40,000	Yes

In this example two inputs one is age and other is salary are two inputs, here for yes condition is (age<66 AND salary>15000). Classification of these two classes is given below



Here black dots represent the one class in which like AND function one condition is false and red dots represent other class in which both conditions are true.

3.2 Multi layer neural network:

It is also known as Multilayer perceptron neural network. It is consist of two layers, three layers. In single layer perceptron we can solve simple problems like linear problem with one decision boundary. But for non-linear problems, we can join two or more single layer perceptrons known as multilayer perceptron. It will classify complex problems, in two layer perceptron neural network classify problem with two decision boundaries and three layer perceptron calculate more complex problems. There are different neural networks. Commom are:

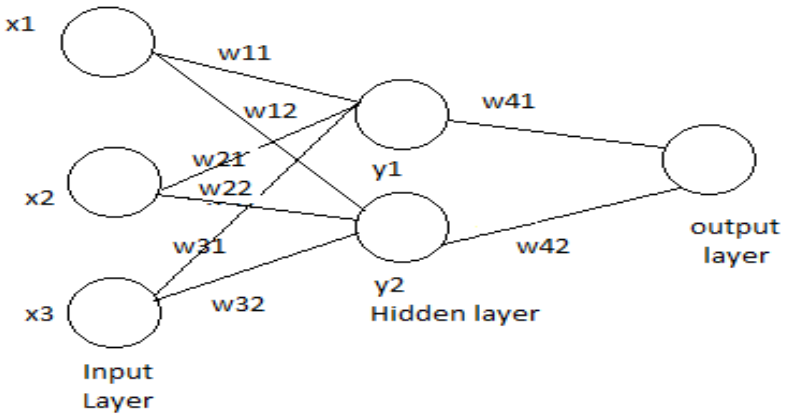
Feedforward (backpropagation)
Associative
Recurrent

Feedforward-backpropagation method: Unlike single layer neural network it consists of three layer input layer, intermediate layer or hidden layer and output layer. Intermediate layer process the output of previous layer.

4. Benefits of hidden layer:

It calculate the complex problems, by choosing several inputs from the previous layer

- Different neuron process at same time, thus it is parallel process.



Here input layer consist of different nodes, each node represent particular attribute. Like age and qualification of person decide eligibility of candidate for particular job. Output layer give answer in terms of 0 or 1, like eligibility of person. Potency of input is decided on base of weight, it is crucial part in neural output of neural network, summation function and transformation function. Summation function is sum of weighted inputs $\sum$ xi*wi and transformational function that compute the activation level neurons. There are different transformational functions one is sigmoidal also known as logarithmic activation that is S-shaped produce output as 1 or Sometimes threshold value can be used, let threshold value is θ and if value more than θ then output will be 1 otherwise it will zero. Sigmoidal function can denoted as $1/1+e^{-y}$ where y is output. It will solve the problem of XOR function that is calculated by Feedforward Neural Network, in this network combination of two neural network done one is AND function Neural Network and other is OR. According to [1]McCulloch pitts method it is represent as below

In this bias b1=-1.55 and bias b2= -.45 and bias b3=-.45 and activation here means activation function θ (), that is one if output more than zero, and zero otherwise.

Neuron 1 acts as AND gate and neuron 2 act as OR gate here and w11=w12=w21=w22=1, truth table for neuron 1 is given below:

X1	W11	X2	W21	X1*W11+X2*W21+bias	Neuron 1
0	1	0	1	0+0-1.55=-1.55<0	0
0	1	1	1	0+1-1.55=-.55<0	0
1	1	0	1	1+0-1.55=-.55<0	0
1	1	1	1	1+1-1.55=.45>0	1

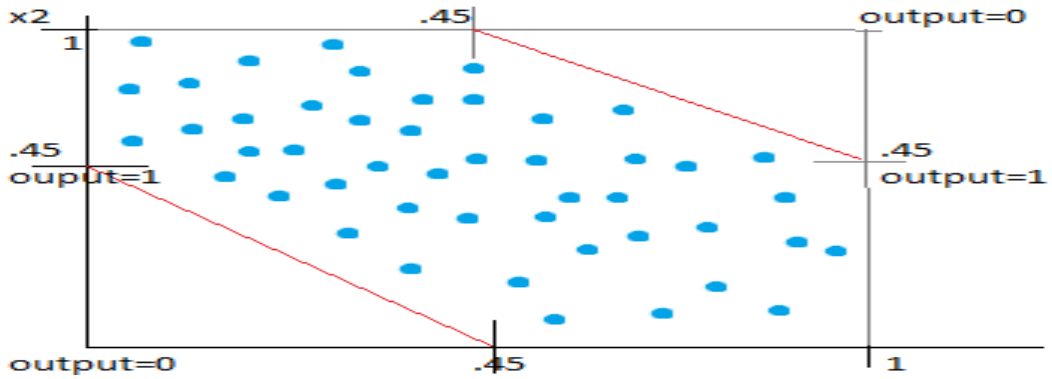
Neuron 2 acts as OR gate represent as below by the truth table below:

X1	W11	X2	W21	X1*W11+X2*W21+bias	Neuron 2
0	1	0	1	0+0-.45=-.45<0	0
0	1	1	1	0+1-.45=-.55>0	1
1	1	0	1	1+0-.45=-.55>0	1
1	1	1	1	1+1-.45=1.55>0	1

Combination of these two tables constructs XOR that is represents below:

X1	X2	N1	W31	N2	W32	N1*W31+N2*W32+BIAS	Neuron 3
0	0	0	-3	0	2	0+0-.45=.45<0	0
0	1	0	-3	1	2	0+2-.45=1.55>0	1
1	0	0	-3	1	2	0+2-.45=1.55>0	1
1	1	1	-3	1	2	-3+2-.45=-1.45<0	0

So in this way multi layer feedforward method construct outputs with two decision boundaries, it is very helpful in data mining where we have to classify on the base of two decision boundaries. Now graphic representation of this graph will below.



Here area with blue dots represent output =1 and area above and below represent output=0

5. Backpropagation algorithm

It is also known as [2] back-error propagation use as supervised learning. It consist more than or equal to two hidden layers. It is simple algorithm to find error, basic steps of Backpropagation algorithm with example is given below. Consider diagram a.1

1. Let training vector= t(1,1,1), associate target=(.2,.5,.2), output from layer 4 and 5=(.5,.5) and activation function used here is logistic function denoted as $f(x)=1/1\text{-}exp^{-x}$

2. First of all calculate the actual output of three nodes
   Node 1=.4*.5+.5*.5=.2+.25=.45 and f (0.45)=$1/1\text{-}exp^{-.45}$=.305
   Node2=.2*.5+.2*.5=.10+.10=.20and f (.20)=$1/1\text{-}exp^{-.2}$=.284
   Node3=.5*.5+.4*.5=.2+.25=.45 and f (.45)= $1/1\text{-}exp^{-.45}$=.305
3. Calculate the error for each output unit
   Err=target-output, here target = (.2, .5, .2)
   First output unit= .2-.305= -.105
   Second output unit=.5-.284= .216
   Third output unit=.2-.305= -.105
4. Error can occur in the hidden layer that can be calculated as
   Err[j] = ∑ i=1to n (w[i]*Err[i])
   Err [4] = -.4*-.105+.2*.216+.5*-.105= -.042+.043-.053= -.052
   Err [5] = .5*-.105-.2*.216-.4*-.105= .053-.0413-.042= .052
5. Now new connection with new weight from 1 to 4 and consider learning rate α= .30
   Calculation of this is done by formula as below
   w[j][i]= w[j][i]+(α* activation[j]*Err[i]*F(e[i])*(1-F(e[i])
   w[4][1]= w[4][1]+(.30*.5*-.105*.305*(1-.305)
   = .4-.003098=.397

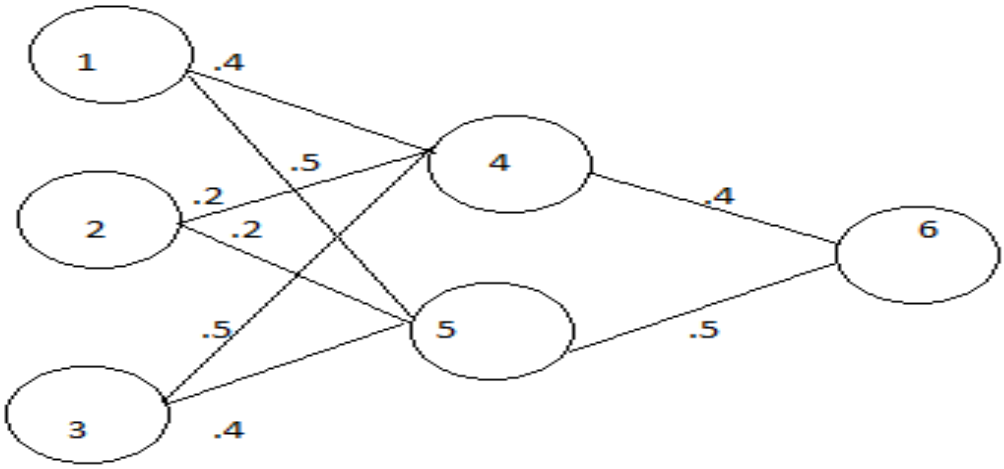
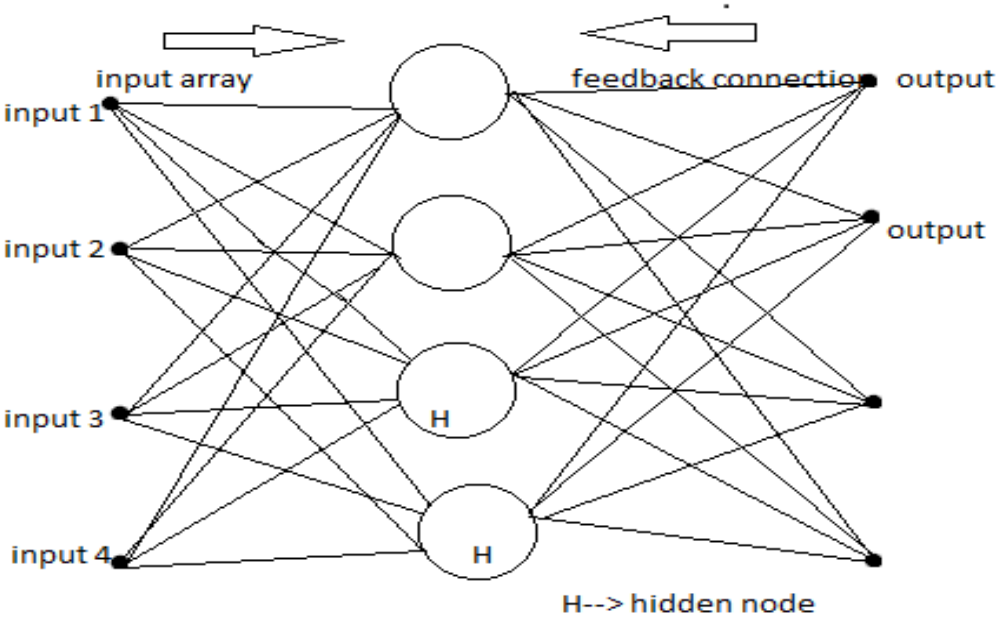


Fig a1

Role of this model in data mining: This model help to solve the static problems like classification and generalization. It also solves the problem of decision boundaries.

6. Recurrent model

This model as compare to single layer and multi-layer neural network help in solve dynamic problems, these problems are time dependent like to make account of process number, sequence number and forecasting marketing data, in the stoke market. This is represents as below:



7. Application of neural network in data mining:

Neural network in data mining identify the fraud detection in tax and credit card. It also detects the bankrupt person. Forecasting is another application by which we can predict the data of future on the base of historic data. Forecasting is done in foreign exchange, stoke market, loan approval, and change in economics. It can also predict the nature of employees in the firm. Customer behavior can also analyze that either customer can spend money.

8. Future scope:

In neural network some drawbacks are there like redundant weights are need to send different data. In brain connection of neurons are approximately equal to 10^{14} and data has to send in less time but in neural network unable to send data in 10^{9} seconds. Data is unlabeled, so these problems can be solves in future in effective way.

9 .Conclusion:

Neural network in data mining plays vital role for classification of the complex data. It helps in generate supervised learning. Algorithms of neural network perform parallel processing. These algorithms are efficient and perform effective result by self adjusting nature.

10. References:

[1] First artificial neurons: The McCulloch-Pitts model By Utkarsh | Published: November 21, 2009

[2] Chapter-6 neural network for data mining

[3] Effective Data MiningUsing Neural Networks by Hongjun Lu, Member, IEEE Computer Society, Rudy Setiono, and Huan Liu, Member, IEEE

[4] Prof. S. Sudarshan CSE Dept, IIT Bombay Most slides courtesy: Prof. Sunita Sarawagi School of IT, IIT Bombay

[5] Research of Data Mining Based on Neural Networks by Xianjun Ni

[6] Artificial Neural Networks and Data Wismar Business SchoolMining by Uwe Lämmel