Breast cancer classification using Neural Network

Nasir Raza

14031282

Neural network and Fuzzy Systems

Course work_1

3/12/2017

1. Introduction

"Breast cancer is a disease in which cells in the breast grow out of control and these cells form a tumor in that area of the body and tumor can be seen through X-ray". [1]

Breast cancer can begin in different parts of the breast. A breast is made up of three main parts: lobules, ducts, and connective tissue. The lobules are the glands that produce milk. The ducts are tubes that carry milk to the nipple. The connective tissue (which consists of fibrous and fatty tissue) surrounds and holds everything together. Most breast cancers begin in the ducts or lobules.

1.1 Types of Breast Cancer:

There are different type of cancers but these two types are the most common

1. **Invasive ductal carcinoma:** In that kind of cancer, cancer cells become outside the ducts into the other parts of the breast tissues.

2. **Invasive lobular carcinoma:** In that type of disease, cancer cells spread outside the lobules to the breast tissues and effect on the surrounding tissues. Invasive cancer cells also effect on the other parts of the body.

When cancer cells become outside the breast through blood veins and lymph vessels and spread to the other parts of the body then it is called metastasized.

1.2 Symptoms:

- New lump in the breast or underarm (armpit).
- Thickening or swelling of part of the breast.
- Irritation or dimpling of breast skin.
- Redness or flaky skin in the nipple area or the breast.
- Pulling in of the nipple or pain in the nipple area.
- Nipple discharge other than breast milk, including blood.
- Any change in the size or the shape of the breast.
- Pain in any area of the breast. [3]

Now a days, this spreading disease generally very common in most of the women. There are many strategies and approaches used to find out that if a woman has breast cancer or not. Likewise, neural networks is also a technique that helps us to find out the disease. We develop a neural network system where we give some data about patient disease to the system and the system tells us that a patient has this disease or not.

3.Background

The first artificial neural network was invented in 1958 by psychologist Frank Rosenblatt. Called Perceptron, it was intended to model how the human brain processed visual data and learned to recognize objects. [4]

Hierachical Neural Network

This is one possible way to organize the processors in a neural net. In this tiered hierarchy, each processor sends its output to all the processors on the next level. One consequence of this is that there's no way to determine the actual path taken by data to produce the final output.

INPUTS

[4]
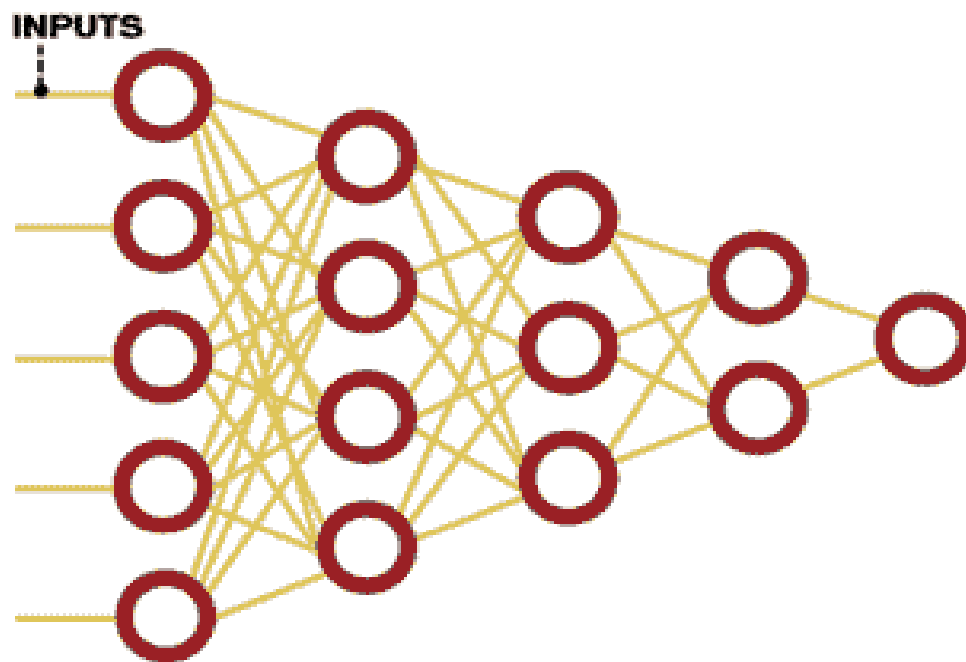
2.1 Neural Network

A neural network typically consists of neurons. A neuron is a basic building block of neural network structure .A neural network consists of layers, there are basically three major layers in a neural network one input layer, one output layer and the other one is hidden layer, hidden layer may consist of one layer or multiple layers. The following figure shows the structure of a simple neural network. [5]

Hidden

Input

Output
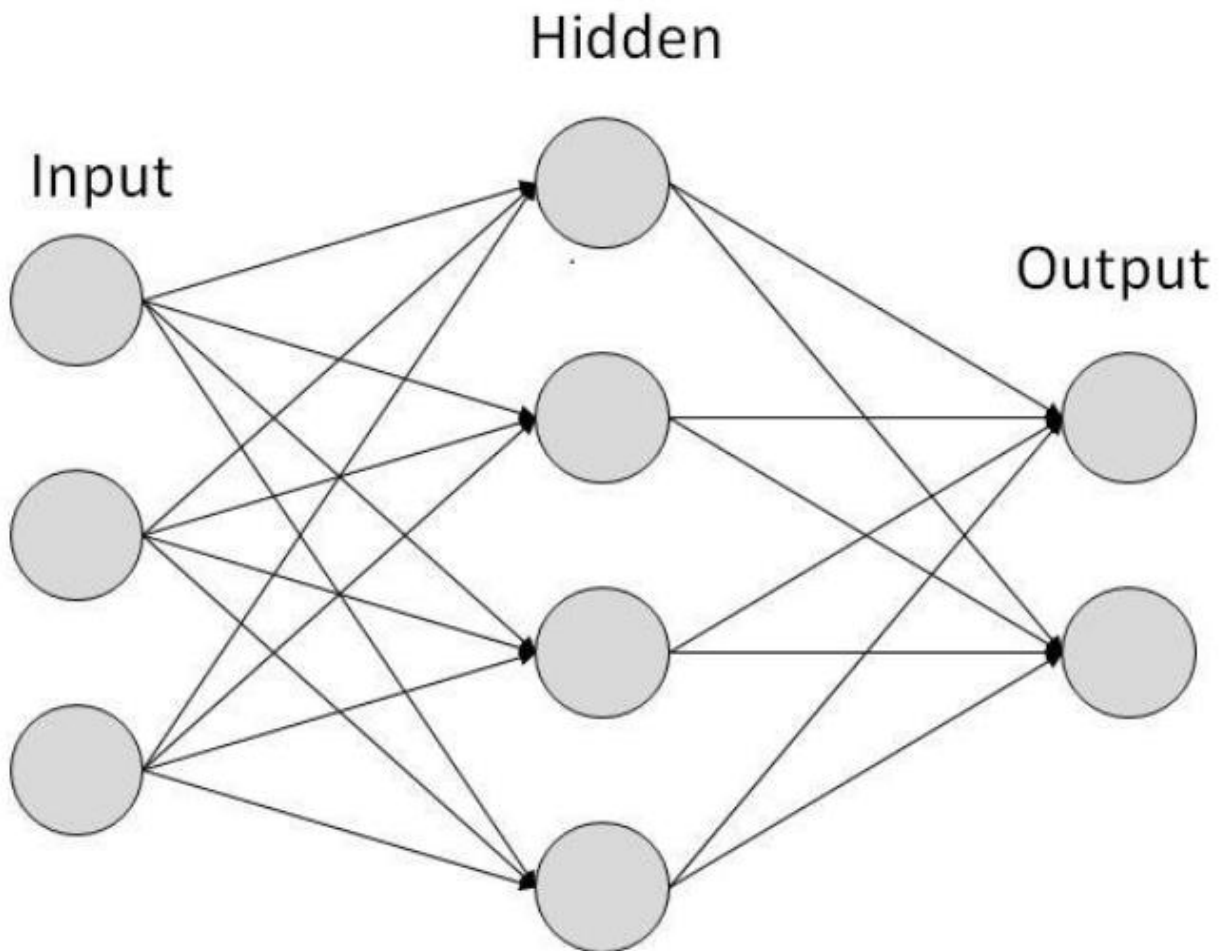
In the following diagram each arrow shows the connection between two neurons and the pathway shows the flow of data stream from one neuron to another neuron. Every connection has

some weights against some numeric value which control the signal between two neurons. A neural network trained against some weights if the network produced some good results in the output then there will be no need to change the weight and if network produce some bad or poor result according to the case   then the system change its weights to output required result.

There are two types of neural networks.

1. Feedforward neural network   **2.** Feedback neural network

3.1.1Feedforward ANN

In Feedforward, data stream is unidirectional. Data send from one unit to another unit from where it does not receive any data information before. These kind of neural network used in pattern generation and classification. They have fixed inputs and outputs and they have not feedback loops.
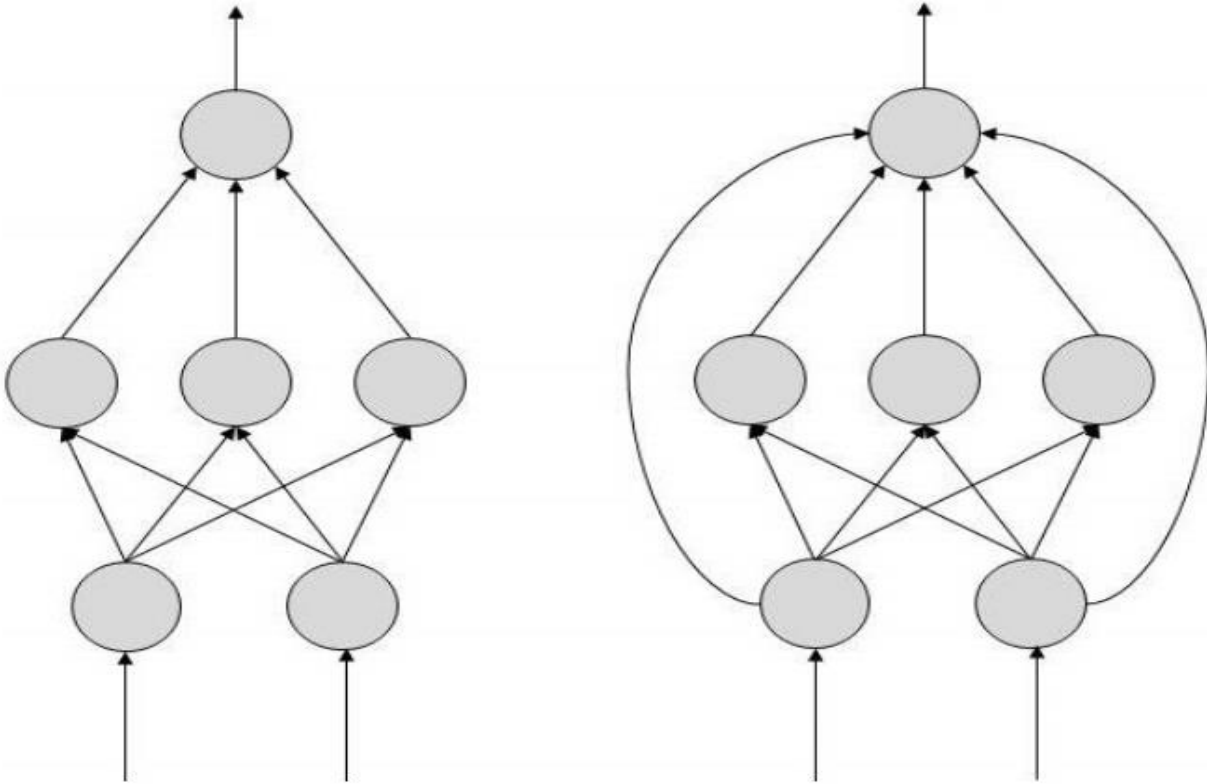
3.1.2Feedback ANN

In Feedback, data stream is bidirectional because feedback loops are allowed. These type of neural network are used in content addressable memories. Following figure shows the structure of Feedback neural network,
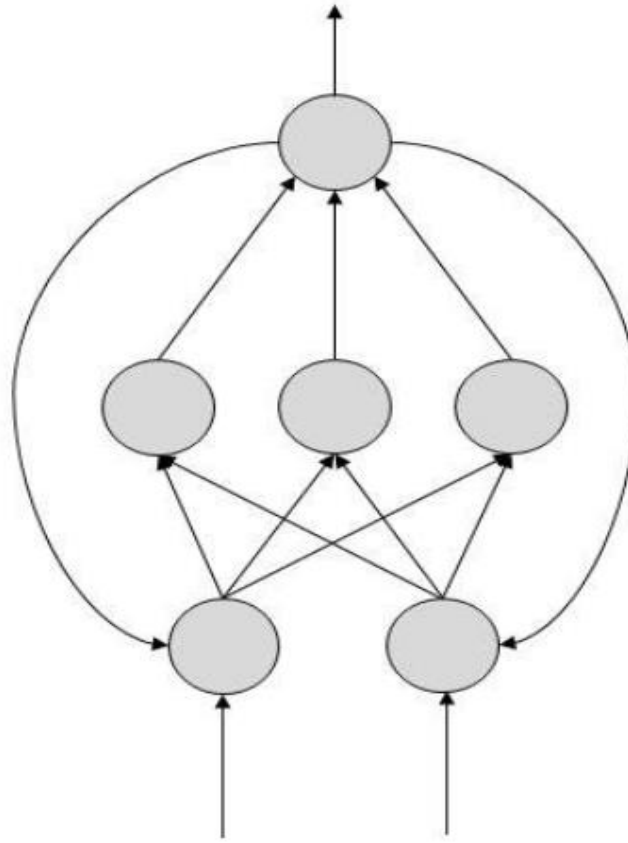
[5]

4. Main Part

This section of the report gives a brief overview about data collection, building neural network system and its methodology.

4.1 DATA COLLECTION:

Firstly, we collect dataset about breast cancer against which we build a neural network from the website UCI machine learning dataset repository. This dataset has total 11 columns, each column has some information against his numeric value. Given down is the information against numeric values which is given in the each column

Attribute Information:

1. Sample code number: id number
   2. Clump Thickness: 1 - 10
   3. Uniformity of Cell Size: 1 - 10
   4. Uniformity of Cell Shape: 1 - 10
   5. Marginal Adhesion: 1 - 10
   6. Single Epithelial Cell Size: 1 - 10
   7. Bare Nuclei: 1 - 10
   8. Bland Chromatin: 1 - 10
   9. Normal Nucleoli: 1 - 10
   10. Mitoses: 1 - 10
   11. Class: (2 for benign, 4 for malignant)

4.2 Load Dataset

There are total 699 cases in dataset, from 16 cases are those which have some missing values against (?).There are different ways to solve these 16 cases but I prefer to neglect these 16 cases and build a neural network system for remaining 683 cases.

4.3 Building Network

After loading the Dataset in Matlab, I build a neural network using this Matlab function newff

net = newff(train_inputData1',train_outputData1',20, {'tansig' 'tansig'}, 'trainr', 'learngd', 'mse');

After creating a neural network for dataset, transfer it to training function with train data to train the network

net = train(net, train_inputData1',train_outputData1');

4.4 Finding Results

After training and testing of the dataset we find the accuracy of the dataset by using this formula

Accuracy= (matching_index / total_ testingdata * 100)

3. Experimental Results and analysis

3.1 Hypothesis 1

My assumption is that if we take training data in increasing order and testing data in decreasing order by 10% of difference margin the accuracy of result should be increase gradually.

Parameters:
goal = 0.01
epochs=6
validation check=100

Result:

 we can see that my assumption is wrong because the accuracy rate in some reading increase by slightly margin and in next reading it decreases and then again increase.so we can say that the by taking increasing training data and decreasing data does not give us as good results as we expected.

Training data	Training data range	Testing data	Testing data range	Accuracy
10%	1-to-68	90%	68-to-683	96.4%
20%	1-to-139	80%	139-to-683	96.3%
30%	1-to-207	70%	207-to-683	97.2%
40%	1-to-275	60%	275-to-683	97.54%
50%	1-to-343	50%	343-to-683	96.4%
60%	1-to-411	40%	411-to-683	98.8%
70%	1-to-479	30%	479-to-683	99.5%
80%	1-to-547	20%	547-to-683	100.7% ,99.2%
90%	1-to-615	10%	615-to-683	100% , 101.1%
100%	1-t0-683	0%	0	Inf

5.2 Hypothesis 2

In this hypothesis, I use same strategy as I used in above hypothesis but I just change the order of training data and testing data in reverse order.

Parameters:

goal = 0.01
epochs=6
validation check=100

Result:

The result was not quite different from above experiment as I expected that the order might effect on the accuracy rate and we get some better result.

Training data	Training data range	Testing data	Testing data range	Accuracy
10%	615-683	90%	1-to-615	92.6%
20%	547-683	80%	1-to-547	93.4%
30%	479-683	70%	1-to-479	93.5%
40%	411-683	60%	1-to-411	95.8%
50%	343-683	50%	1-to-343	95.02%
60%	275-683	40%	1-to-275	93.4%
70%	207-683	30%	1-to-207	92.2%
80%	139-683	20%	1-to-139	93.4%
90%	68-683	10%	1-to-68	89.5%
100%	1-683	0%	0	100

5.3 Hypothesis 3

I assume that by increasing the value of epochs the accuracy rate should ideally increase.

Parameters:

goal = 0.01
epochs=6
Validation check=100
Training Data = 50%
Testing Data=50%

Result:
We see that the result was surprising because it was not that as we expect.

epochs	Accuracy
6	98.5%
8	98.2%
10	99.1%
12	98.2%

6. Bibliography and Citations:

[1]"CDC - What Is Breast Cancer?", *Cdc.gov*, 2017. [Online]. Available: https://www.cdc.gov/cancer/breast/basic_info/what-is-breast-cancer.htm. [Accessed: 03- Dec- 2017].

[3]"CDC - What Are the Symptoms of Breast Cancer?", *Cdc.gov*, 2017. [Online]. Available: https://www.cdc.gov/cancer/breast/basic_info/symptoms.htm. [Accessed: 03- Dec- 2017].

[4]A. Kay, "Artificial Neural Networks", *Computerworld*, 2017. [Online]. Available: https://www.computerworld.com/article/2591759/app-development/artificial-neural-networks.html. [Accessed: 03- Dec- 2017]

[5]"Artificial Intelligence Neural Networks", *www.tutorialspoint.com*, 2017. [Online]. Available: https://www.tutorialspoint.com/artificial_intelligence/artificial_intelligence_neural_networks.htm. [Accessed: 03- Dec- 2017].

[6]"UCI Machine Learning Repository: Breast Cancer Wisconsin (Original) Data Set", *Archive.ics.uci.edu*, 2017. [Online]. Available: http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Original%29. [Accessed: 03- Dec- 2017].