Neural Network and Fuzzy Systems
Coursework – 1

Breast Cancer Diagnostics with the help of Neural Networks.

Usama Saleem

14031202

Final Year

Abstract:

Self-Driving cars, SIRI, Cortana, Robots and many more are nothing but mere wonders of Artificial Intelligence. It has been increasing rapidly and swiftly and has proved out to be a huge innovation in the field of Computer Science. Artificial Intelligence today is usually referred as AI, is been designed to perform a lot of tasks, like Facial Recognition or Internet searches but the long-term aim and motivation of many researchers is to create a General AI or strong AI that will be able to perform all the tasks. It may outperform humans as well. Already AI is being used in Chess, and in equation solving. Among AI's many and numerous focuses, we aim to learn its help and implementation in disease diagnostics and classifications. The particular type of disease we will be focusing on is the Breast Cancer, for which a particular dataset has been provided to us. Our task is to train a Neural Network that will be able to classify if the cancer can be diagnosed or cured or not.

Introduction:

Breast Cancer is the most well-known intrusive cancer in women, and the second primary cause of malignancy death in ladies, after lung disease. According to recent studies and researches, it has been discovered that, the probability of any woman dying from Breast Cancer is **2.7%**. Which means that, among every 37 woman there is likely a chance of 1 woman who can die due to breast cancer. Researchers and doctors have found many ways to tackle this disease, and among all those solutions, one solution is to diagnose the breast cancer is through the **Neural Networks** so that the disease is fully diagnosed and completely treated. **[2]**

Also, there are two different states of Breast Cancer, which are following,

Malignant:

If the diseased cells are allowed to grow in the nearby tissues and are allowed to spread in the whole body, then those diseased cells can lead to patient's death.

Benign:

The cells do not spread out of the breast; rather it is the abnormal growth of the cells in the breast. That means when the state of cancer is benign then it can be cured and not as lethal as the Malignant state.

Background:

The Neural Network is a collection of different neurons of the brain. Neurons can be merged together to form the neural network just like Human's brain. The neurons are organized and interconnected, creating a very complex structure and are responsible for sending the signals. Now the magnitude of these signals can be increased and decreased towards any direction. Inside the neuron, there is an activation function that decides that whether to forward a particular signal to the other neuron or not. There are different layers in a neuron, these neurons are arranged in some particular layers and the connection between the layers is formed. The first layer receives the input and is eventually called the **Input Layer**. The last layer is responsible for sending the output so it is called the **Output Layer**. Between these two layers, another layer exists which is called the **Hidden Layer**. The basic structure of the Neural Network is shown in the figure below:

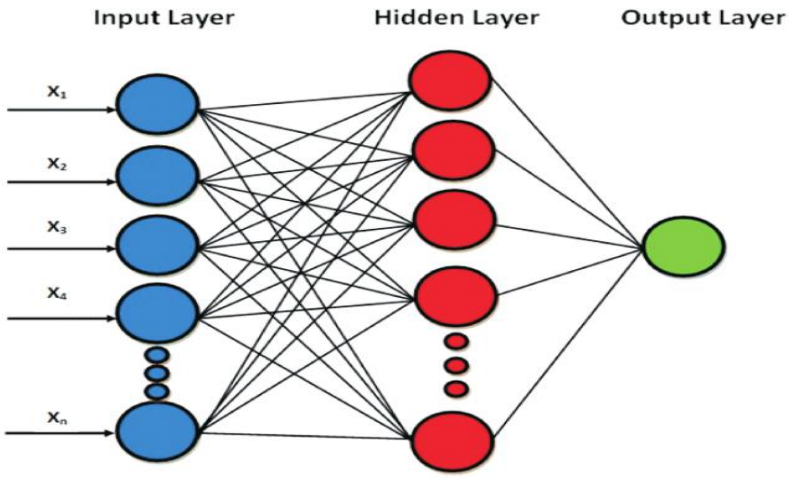

Figure 1: Artificial Neural Network

As it can be seen in the above picture, Multiple inputs are given to the neural network like x1, x2 etc. These nodes or weights can be altered after every input and then the mesh of neurons in the neural network, form a complex structure as it can be clearly seen in the figure 1. The Neurons, as mentioned earlier have an activation function that helps them in deciding that whether to forward a particular signal to the other or not. A trained neural network is expert to categorize the different dataset given to it. From the given dataset, it has the ability to detect the complex patterns and that are only identified by the higher computer techniques. Its training can be done by changing the weights of the neurons so that it can categorize the data precisely.

There are two types of Artificial Neural Network,

Feed Forward Network:

In this type of network, the neural signals can only travel in one way that is, from input to output. This type of Artificial Neural Network can be used in pattern recognition.

Feedback Network:

It allows the signals to travel back-side as well by using loops. That is why it is more powerful and useful than feed-forward.

Previously Work Done by Various Organizations:

Different Organizations have been working on the awareness and diagnosis to kill the breast cancer. Artificial Neural Networks, because of its rapid action and identification, have been used widely to solve many complex problems. Salim et al, have already developed the breast cancer diagnosis by using the Artificial Neural Network and hybrid magnetoacoustics method (HMM). Pradhan and Sahu have also foretold the type of breast cancer by using the soft computing.

Different solutions currently exist in different programming and computing techniques and languages like MATLAB and Java etc. Kala et al, have implemented the Artificial Neural Network and its performance and accuracy was **96%** on the training data and **95%** on the testing data. **[4]**

Training of Neural Network & Methodology:

University of Wisconsin Hospitals have provided the dataset. This dataset is in the text form and to use it in the MATLAB we have converted the text file into **'.m'** file.

As specified, the primary column of the dataset is the patient ID, which is no utilization of preparing the network so we remove expel the first column from the dataset. The last column of the dataset is the output against every evaluation figure. It is valuable for later utilize yet to train, it isn't helpful as it might influence the network. Therefore, we need to delete the first and last columns of the dataset to appropriately prepare the network. In the wake of deleting the columns, there is another problem that the dataset contains question marks **'?'** and for preparing the neural network the question mark isn't useful. Subsequently, we need to transform it to **'0'.**

Additionally, the output or the last column in the dataset is either '2' or '4' i.e. 2 denoting the **'Benign'** and 4 representing **'Malignant'**. Benign condition of cancer is treatable however the threatening state is

Malignant and we need to characterize these states. As we are using the **'tansig'** function, which has its range from - 1 to 1 hence, we need to change over the output framework into either - 1 to 1 or 0 to 1 shape. This will turn out to be valuable for our outcomes. At last, we will separate the dataset into two sections i.e. one for training and the other for testing.

Hypothesis:

Firstly, I assume that if I provide more amount of training data, then the neural network will give the most accurate and precise results. Because as the neural network and neurons work similarly like the Human's brain so it will have more data to work. Therefore, for the first hypothesis and experiment I am going to divide the data into the proposition of 60 and 40. That is 60% trainingData and 40% TestingData and we will see the results of the experiment with this hypothesis. The code is attached in the Matlab file with the extension '.m'.

Result:

The result was in accordance with the hypothesis, that the higher the training data, the more accurate the result will be. With the data given by me to the neural network in the first hypothesis, the results were good and quite accurate. There was also a lot of probability of the errors with the current hypothesis which I tried to minimize in the next hypothesis. The results of first hypothesis are given below:

trainingData	**testingData**	**Accuracy**	**Error**
60	40	77.2501	22.7499

Hypothesis:

As already mentioned in the previous hypothesis, the trainingData amount was slightly higher than the amount of the testingData, so in my next experiment I chose the different set of values from the dataset given, and increased the amount of testingData by 10% and that resulted in the decrease in the amount of the testingData by 10% alse. So, the total data distribution become 70-30.

Result:

The hypothesis was fine and the result was very accurate, which shows that with the set of values taken from the dataset, and the data distribution made, helped in triggering the accuracy of the neural network to **99%**. The accuracy and error figure are shown below:

trainingData	**testingData**	**Accuracy**	**Error**
70	30	99.0743	0.9257

Hypothesis:

As the assumption made before starting the experiments, has worked quite well so far, so I constantly increased the data distribution of the trainingData and the TestingData by 10%. And checked the output and noted the accuracy and the error percentages.

Result:

This time, the data distribution made in the above hypothesis showed that the accuracy percentages has slightly decreased than the experiment conducted with the previous hypothesis. This shows that this data distribution and the set of values taken from the dataset was not as accurate as the values taken from the previous set of datasets.

trainingData	**testingData**	**Accuracy**	**Error**
80	20	97.0317	2.9863

Hypothesis:

In my last hypothesis, I again increased the data distribution among the trainingData and the testingData by 10% and then finally noted the accuracy and the error percentages.

Results:

This time again, the accuracy was almost as same as the 70-30 data distribution, which also made the hypothesis true. And then I noted the accuracy and the error percentages that are shown below:

trainingData	**testingData**	**Accuracy**	**Error**
90	10	99.0743	0.9257

Conclusion:

As mentioned earlier in the hypothesis and assumption, that the higher the value of the trainingData will be the higher will be the accuracy and less error probability. We get the maximum result when the distribution is 70-30 and 90-10. Also, when the distribution is 60-40 the accuracy drops and the error percentage increases gradually. The reason for both the distribution to be same lies in the fact the how the initial weights are assigned and how the values are taken from the dataset given to us. The overall data distribution and accuracy table is given below:

trainingData	**testingData**	**Accuracy**	**Error**
90	10	99.0743	0.9257
80	20	97.0317	2.9863
70	30	99.0743	0.9257
60	40	77.2501	22.7499

References and Bibliography:

[1] *"UCI Machine Learning Repository – RP"*, Available At: http://www.researchpipeline.com/mediawiki/index.php?title=UCI_Machine_Learn ing_Repository , (Accessed Date: 08 December, 2016)

[2] A. Ahmad, (2012), *"Probabilistic neural network for breast cancer classification"*, Available At: http://link.springer.com/article/10.1007/s00521-0121134-8 , (Accessed Date: 01 December, 2017)

https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/

https://www.medicalnewstoday.com/articles/37136.php