Breast cancer detection using neural networks

Abdul Rafay

13031168

Neural Network and Fuzzy System

Course work 1

Course instructor: Dr. Junaid Akhtar

Contents

Abstract:

The main purpose of designing such system is to replace or decrease the human interaction and making the software self-dependent for the identification of “breast cancer” which is an open study area nowadays and it can be identified using these technologies without human involvement. One of the important and intelligent tool for automation is “Artificial neural networks”. The network is given an amount of training data initially which the network uses to train itself and understand the problem and get to the possible solution by itself. Then the network is given the different set of data to operate. The result that comes from the system is derived from the data previously learned by the system. Different types of experiment were done on this system and the best result on a new case is somewhere between 97-99%.

Introduction:

“Breast cancer affects one in eight women during their lives” **[2].** “Breast cancer is the most common cancer in women worldwide. Approximately 1.7 million new cases were recorded globally in 2012, accounting for 25 per cent of all new cases of cancer in women. It is the fifth most common cause of death from cancer in women”[1]. ~~Many efforts and research have been made in this field by keeping the fact and figures to solve this as it is a big problem nowadays for our women and future~~. Nowadays the most widely researched and proposed solution for the detection of breast cancer is by using neural network. The main purpose to use this technique is that it is much faster in generating the result and in diagnosing the disease in the breast cancer patients based on symptoms. The main purpose of writing this report is describe that how neural networks are used to classify cases of breast cancer in women using back propagation, which includes the designing of the network and to deduce the result obtained from the network.

Background:

Neural network architecture is designed by keeping in mind the structure of human mind. The inventor of the first neurocomputer, Dr. Robert Hecht-Nielsen, defines a neural network as –"a computing system made up of a number of simple, highly interconnected processing elements, which process information by their dynamic state response to external inputs”**[3].** An artificial neuron has many inputs but one output. The inputs are passed to a transfer function which gives an output. There are different types of activation functions which can be used in neuron. The type of the function to be used depends upon the person going to develop it. In this way the different numbers of such neurons are used together to create a neural network.

Preprocessing:

The dataset ~~which is used~~ for our experiments is Wisconsin Diagnostic Breast Cancer (WBDC) dataset. The dataset consist of approximately seven hundred breast cancer cases with their result i.e. either benign or malignant. The dataset is divided into two portions i.e. training

data and test data. The training data is used to train the neural network and the test data is used to test that neural network and give the result of our experiment on the basis of trained network. The dataset consist of eleven columns and six hundred and ninety nine rows. First columns consist of "Sample Code Number" which is the id number of patient cases. Next nine columns consist of disease symptoms and the last column consist of class which is either 2 or 4 (2 for benign and 4 for malignant). The input data consist of columns 2-10. Last column is separated and termed as an output for training the network. In this dataset there are 16 instances in column six which has the value **"?"**. This value is firstly replaced by "0" by using "Find and replace" command and then replaced by the mean value of the values in the six column. The platform used for creating and testing the network is **"Matlab"**. The function used to create the network is known as **"newff".** Preprocessing is that I separate the output values of training data and testing data. All the attributes are set like learning rate, goals, epochs and activation function.

Network Architecture:

After all the preprocessing is done and our network architecture is built and then it is sent with training data and function in order to train the network. Then is sent to **"SIM function"** of MATLAB with output values of training class and testing class to compare them and give the result. After training the function measures the accuracy of the network. The formula used by me to calculate the accuracy of the network is:

PercentageAccuracy= ((size (testClass)-error)/size (testClass))*100;

Experimental Results and analysis:

Firstly certain hypothesis were made about different perimeter and then the results were calculated to check that either the hypothesis made by me were true or false. This was all done with order and all the steps were repeated for every experiment. Following are some hypothesis which were verified by experimental analysis.

Hypothesis 1:

Hypothesis statement: Increasing the training data will increase the accuracy of the network. If we give more training data to the network, the network will become more trained and then after testing it may give more accurate result.

- Effects of data distribution on the accuracy of network

Training Data (%)	Testing Data (%)	Accuracy (%)	Error (%)
70	30	99.59	0.40

60	40	98.57	1.42
50	50	98.28	1.71
40	60	97.15	2.85

Result: From above table we can see that if we increase the training data gradually it will result in increase of accuracy. So the hypothesis is true. The reason behind this result is that if we increase the training data it will result in improvement of network, thus result in increase of accuracy and reduction of error.

Hypothesis 2:

Hypothesis statement: Increasing the learning rate will decrease the accuracy and eventually increase in error. The learning rate defines how the fast or slow the network trains on the provided input data. Another main reason which may affect this experiment is number of factors for each input. More factors means the learning rate needs to be lower.

- Effects of learning rate on accuracy and error

Learning Rate	Accuracy	Error
0.05	98.0127	1.9873
0.04	98.9251	1.0749
0.03	99.3149	0.6851
0.02	99.5512	0.4488

Result: From above table we can see that increasing the learning rate will decrease the accuracy. So, we can say that our hypothesis is true.

Hypothesis 3:

Hypothesis statement: Using the proper activation function will increase the accuracy of the network. As I am using **"tansig"** function so it may give more accurate result then other activation function. If we use tansig function on both the hidden layer and the output layer, we should get maximum accuracy than using any other combination of activation functions.

- Effect of change of transfer functions on accuracy and error

Activation Functions (hidden layer, output layer)	Accuracy (%)	Error (%)
logsig, logsig	94.28	5.71
logsig, tansig	96.28	3.71
tansig, logsig	96.85	3.14
tansig, tansig	97.42	2.57

Result: From above table we can see that as we predicted that if we use the **"tansig function"** for both the hidden and output layer we can get the more accurate results.

Conclusion:

Neural networks are widely being used nowadays for the identification and classification of breast cancer. ~~A lot of study is being done on breast cancer using this technique because it is most precise and time consuming process for identification of the problem.~~ There are many parameters which can be tested and improved to increase the accuracy of the network but we have been focusing on hidden layers weights and data distribution. After all these experiment I had done on this network we can see that if we work upon some of the parameters we can increase the accuracy as well as the time consumption as nowadays our big problem is time. We are required to work upon this research as it is a big problem nowadays for our women.

References:

[1]: Diet, nutrition, physical activity and breast cancer. [ONLINE] Available at: http://www.wcrf.org/sites/default/files/CUP_BREAST_REPORT_2017_WEB.pdf [Accessed 29 November 2017]

[2]: Breast Cancer: MedlinePlus. 2016. Breast Cancer: MedlinePlus. [ONLINE] Available at: https://medlineplus.gov/breastcancer.html. [Accessed 29 November 2017].

[3]: tutorialspoint.com. 2016. Artificial Intelligence Neural Networks. [ONLINE] Available at: https://www.tutorialspoint.com/artificial_intelligence/artificial_intelligence_neural_networks.htm. [Accessed 29 November 2017].