

**Breast Cancer Detection**

Muhammad Usman Ghani

14031259

Instructor Dr Junaid

Neural Network and Fuzzy System

Course Work 1

**Abstract:**

In this report a neural network is used to classify whether someone has a breast cancer or not on the basis of given data. We are provided with small amount of cases which will be used to train a neural network and then this neural network will give us results about cases which are not seen by this very network.

**Introduction:**

In recent work and studies neural network has been extensively used in various fields. For example these are used in artificial intelligence, machine learning, medical-diagnosis and so on. In this coursework neural network is used for the diagnostic of breast cancer. There are many different methods of detecting this disease but we are using neural network. We are given some data which will help to classify whether someone has breast cancer or not. Recently neural networks has a popular tool in the classification of cancers, in our case it is breast cancer [1]. Classification is one the most active research in neural networks. The process of classification involves assigning objects into predefined groups based on the number of observed attributes related to those objects and there are many different tools for it.

**Background:**

Breast cancer is the disease which today affects more women in the world. Breast cancer is the second leading cause of cancer deaths worldwide and occurs in one out of eight women [2]. There are many different techniques for its detection for example one is medical imaging technique which is used to detect breast cancer [3].The other is breast screening **“**Breast screening is a method of detecting breast cancer at a very early stage. The first step involves taking an X-ray, called a mammogram, of each breast. The mammogram can detect small changes in breast tissue which may indicate cancers which are too small to be felt.” [4]. One more method is an evolutionary artificial intelligence neural network approachbased on the pareto-differential evolution (PDE) algorithm augmented with local search for the prediction of breast cancer. The approach is named memetic pareto artificial neural network (MPANN) [5]. Technique which we are using is to train a neural network. There is given data which has some attributes which helps classify if someone has malignant cancer or benign cancer. In this coursework we used a neural network for its classification to understand deep learning of how a neural network work and what its future uses are.

The structure of neural network is just like human brain. A neuron is also call a node. It is composed of large number of highly interconnected neurons working to solve a given problem. Neural network classify objects in simple words it takes data as input derive rules based on those data and make decisions. First a neural network is trained on a specific data and then it learns on that given data and then at end we can test unseen data on it, just like human brain. Neural network are used to detect trends and extract patterns which are too complex for human naked eye or by other computer systems. A neural network has some capabilities e.g. Adaptive learning, Self-Organization, Real time operation etc.

**Main Part:**

**Data Description**

* Total Instances = 699
* Total Attributes per instance = 11
* Missing Attributes = 16
* Benign Cases (2) = 458 (65.5%)
* Malignant Cases (4) = 241 (34.5%)
* Column 1 = Patient ID
* Column 2-10 = Input Data (Attribute Information)
* Column 11 = Output Data (Class Attributes)

**Attributes Information**

Columns (2-9) in the data set possess following information:

Sample code number

* Clump Thickness
* Uniformity of Cell Size
* Uniformity of Cell Shape
* Marginal Adhesion
* Single Epithelial Cell Size
* Bare Nuclei
* Bland Chromatin
* Normal Nucleoli
* Mitoses

This section contains data gathering and how a neural network is trained.

**Pre-processing and missing values:**

We used dataset which was given to us. Data had 11 columns and 1st columns was id of patient so we didn’t used 1st column for neither training nor testing in our overall process. Columns two to ten (2-10) were actual attributes which would help us in detecting whether someone has malignant or benign cancer. Last column 11 were the values either 2 or 4 determining benign or malignant respectively. There were also some missing values which were replaced by “?”. As all these “?” were present in column 7 so I manually took mean of all column 7 and result was 3 so I replaced “?” with value 3. I could have simply deleted those cases or would have replaced “?” with 0 but I chose this mean method for missing values.

**Training and testing a network:**

After all data was in correct form and it was loaded in matlab we created a neural network and set its attributes e.g. goal, epochs etc. when training function was built we built testing function on unseen data. After training and testing I wrote a method of finding accuracy in percantage.

**Experimental Results and Analysis:**

In order to properly understand the performance and working of the artificial neural network, various hypothesis has been made and then on the base of these hypothesis, various experiments by changing the parameters of the neural network has been made. Therefore network was trained and simulated on various proportions of the input and testing data. The performance of the proposed system was calculated in terms of accuracy while keeping the number of iterations same for each experiment. The experimental results and brief analysis is provided below:

**Role of Data Proportion**

**Hypothesis:**

If we increase training data accuracy should be increased because large training data represents more diversity and hence its accuracy should be increased by increasing training data. I increased number of training cases by 10% in every experiment and tested a trained neural network on unseen data. So I tried different number of experiments on the given data and following were the results.

**Experiments:**

Thus following is the experiment based on this hypothesis.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Trainingdata** | **Datarange** | **Testingdata** | **Datarange** | **Accuracy** |
| 20 | 1-140 | 80% | 141-699 | 94.41 |
| 30 | 1-210 | 70% | 211-699 | 94.47 |
| 40 | 1-280 | 60% | 281-699 | 95.94 |
| 50 | 1-350 | 50% | 351-699 | 95.98 |
| 60 | 1-420 | 40% | 421-699 | 98.20 |

**Conclusion:**

As my hypothesis was that increase in number of training data will result in better accuracy so hypothesis were true as we can see that accuracy is increased as we increased our training data. The reason is that when we use more data in the training process, then our network get training on a lot of cases. So, when it is run on seen sample, it shows good results, because it has already got training on relevant data examples. So our overall performance and accuracy of the artificial neural network will increase if we use more data in the training.

**Role of Hidden Layers**

**Hypothesis:**

There are hidden layers in neural network, a hidden layer is a neuron which is connected to the inputs of the other neurons and is therefore called hidden layer. There are specific number of neurons on each hidden layer and we can alternate these neurons. Our input data is mapped with these number of neurons. In this case number of neurons twenty.

So hypothesis is that if we increase number of hidden layers by not increasing neurons in each layer our accuracy should be increased because learning time of a neural network will be increased and hence its performance will also be increased.

**Experiments:**

So I tried this experiment by keeping everything constant except number of hidden layers. I gradually increased number of hidden layers by one. Everything else was constant including training data, testing data and number of neurons in every hidden layer.

|  |  |  |
| --- | --- | --- |
| **Number of hidden layers** | **Number of neurons in each hidden layer** | **Accuracy** |
| 1 | 20 | 91.89 |
| 2 | 20 | 92.02 |
| 3 | 20 | 92.63 |
| 4 | 20 | 94.27 |
| 5 | 20 | 91.27 |

**Conclusion:**

As we can see from the above table that as we increased number of hidden layers our accuracy is increased so our hypothesis was true. The reason is that when we use more hidden layers, our network get a lot of spaces to map the weights from input layers. So, by increasing the hidden layer, our performance and accuracy of the network increases to great extent.

**Role of No of Epochs in Training**

**Hypothesis:**

If we increase number of epochs training time will become longer hence its accuracy should be increased. Epochs are basically number of iteration on which our neural network should be trained.

**Experiments:**

|  |  |
| --- | --- |
| **No of epochs** | **Accuracy** |
| 50 | 95.69 |
| 100 | 95.28 |
| 150 | 94.88 |
| 200 | 94.68 |
| 250 | 93.04 |

**Conclusion:**

As we can see from the above experiment that by increasing the number of epochs (which is number of iterations) our accuracy is gradually decreased. So it is concluded that too much number of iterations can also decrease accuracy. The reason is that data has already got training in few epochs. So, increase in epochs will not increase the accuracy, rather it will have adverse effect on the accuracy of the artificial neural network.

**References:**

|  |
| --- |
|  |
| 1.Arun George Eapen, master thesis, Application of Data Mining in Medical Applications, Waterloo,  Ontario, Canada, 2004.  2. Azar, A.T. and El-Said, S.A., 2013. Probabilistic neural network for breast cancer classification. *Neural Computing and Applications*, *23*(6), pp.1737-1751.  3. Mehdy, M.M., Ng, P.Y., Shair, E.F., Saleh, N.I. and Gomes, C., 2017. Artificial Neural Networks in Image Processing for Early Detection of Breast Cancer. *Computational and Mathematical Methods in Medicine*, *2017*. |
| Vancouver |  |

4. Menéndez, L.Á., de Cos Juez, F.J., Lasheras, F.S. and Riesgo, J.Á., 2010. Artificial neural networks applied to cancer detection in a breast screening programme. *Mathematical and Computer Modelling*, *52*(7), pp.983-991.

5. Abbass, H.A., 2002. An evolutionary artificial neural networks approach for breast cancer diagnosis. *Artificial intelligence in Medicine*, *25*(3), pp.265-281.