2017

Breast Cancer Classification Using Neural Networks

Hassan Ahmed

14031239

Submitted to: Dr. Junaid Akhtar

Summary

Artificial Neural Networks are statistical learning models which have amazing abilities to get useful information from raw and indefinite data. Therefore the neural network can be used to figure out various trends from input data, which are quite complicated for humans.

In this analysis, artificial neural network (ANN) has been used to make decision support system (DSS) to classify the severity of breast cancer (as benign or malignant) based on statistical input data, contained 699 instances.

Introduction

In spite of a lot of advancement in medical science, breast cancer is still fatal disease. Breast cancer is the most prevalent type of cancer in women and is one of the major causes of cancer morality in women aged 20-59 [1]. Depending upon the potential harm of the breast cancer, it has been classified into benign and malignant type [1].

To address that issue, various techniques of data mining and artificial intelligence like artificial neural network, decision tree, logistic regression and genetic algorithms were used [2].

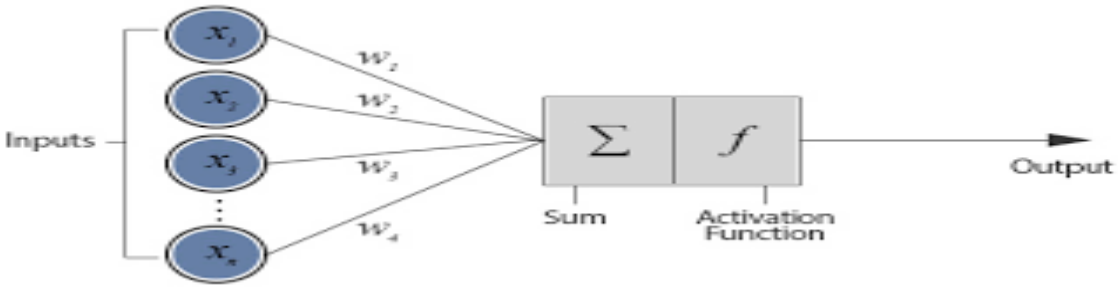
Artificial neural (ANN) is one of the best artificial intelligence techniques for common data mining tasks, such as classification and regression problems. Research showed that ANN delivered good accuracy in breast cancer diagnosis [1]. That report also shows a thorough analysis to make such network by learning from data of breast cancer and making wise decision on unseen samples.

Background

Artificial Neural Network Model

Artificial Neurons

Neural network is mostly contained on individual or interconnected models, usually called neurons or nodes. The neurons are basic building blocks of the neural network. A set of sample input(s) are given to neuron, then these inputs are forwarded to a decision making function, which is called activation function. Then that activation function generates the output based on given input.

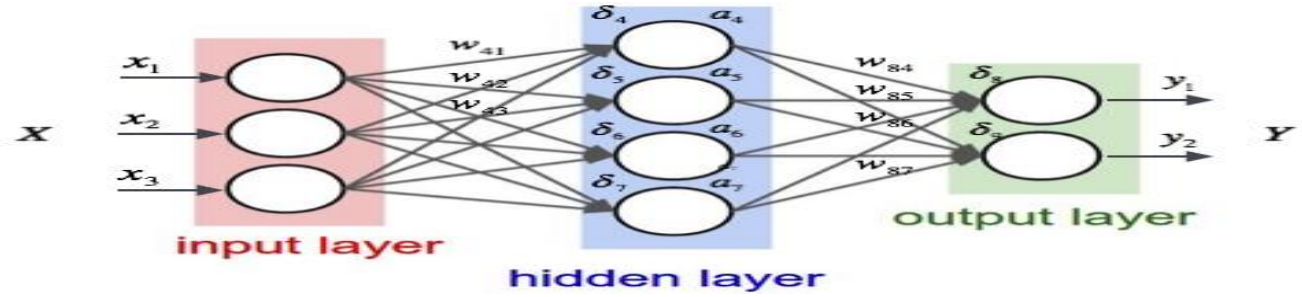


[3]

Neural Network Structure

Neural network is compromised on three types of layers which are highly interconnected with neurons:

- Input layer
- Hidden layer(s)
- Output layer



[4]

The sample inputs are presented to neural network through input layer, which communicates with one or many hidden layers, where the main processing is done. These hidden layers are linked with output layer, where the original results are shown.

Previous Work on Breast Cancer

For past some years scientist and doctors are trying their best to find suitable solution to detect breast cancer at early stages through various techniques. There are about 1.38 million new cases and 458 000 deaths from breast cancer each year [5].

A research was done by Dr Medhat and Mushammad Farouq about the abilities of support vector machine (SVM) classifications with tree boost and tree forest in analysing the dataset for the extraction of mass features that discriminates the true and false cases. Here SV showed better results as compared to those of tree boost and tree forest [6].

Satisfactory work has also been done by Wei-pin Chang and Liou. They explored that genetic algorithm techniques yielded better results than other data mining techniques in order to analyse the data of breast cancer in terms of overall accuracy of the patient classification [6].

Moreover, K. Rajiv Gandhi, Marcus Karnan and S. Kanan used Particle Swarm Optimization Algorithm for the classifications of the breast cancer datasets [6].

Senapati used linear wavelet neural network and '*Firefly Algorithm*' for further improvement to find the solution of Breast cancer using *Wisconsin Diagnostic Breast Cancer* dataset. From their study they successfully classified the dataset for benign and malignant cases with accuracy rate more than 90% for both cases [7].

Main Part

In that task we have been given a sample data consist on 699 instances. That data is about the breast cancer. On the base of that sample data input the severity of breast cancer can be categorized as benign or malignant.

Description about Dataset

The major details about the dataset which has been collected from Wisconsin Breast Cancer Database (WBCD) was analysed and following information figured out [8] [9]:

Dataset Description

After analysing the data, following information were found [10]:

- Total no's of samples = 699
- Total columns per row (data is in matrix form) = 11
- Missing Attributes = 16 (These values are not available in the given dataset)
- Benign Cases (which has been indicated as 2) = 458 (65.5%)
- Malignant Cases (which has been indicated as 4) = 241 (34.5%)
- Column 1 = Id Number
- Column 2-10 = Sample Input Data
- Column 11 = Output Data (Class Attributes)

Details about Each Attribute of Row

Each row consist on 11 columns. The details about each column value is following [10] :

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)

Neural Network Formation

Data Processing

The given dataset for the breast cancer is in raw form. In order to use that dataset in the training of neural network, it is quite necessary to arrange the dataset in proper way. The first column is consist on the Id Number, which is quite unnecessary in the training of neural network. So that column will be skipped.

There are some missing values (indicated with "?"), so we will take mean of the all values of the column(s) in which the values are missing, in order to find out the suitable alternate for missing values.

While the time of data loading we manually, replaced "?" with 0. Then that 0 was replaced with 3. Because the mean of col 7 (in which values were missing) is 3.

```
mean(inputData(: , 7))
inputData (inputData == 0) = 3;
```

After pre-processing we placed the refined dataset in the variable named as DATA, which we used in our code.

Components Involved in the Training of Neural Network

In order to generate the desirous outputs based on given sample input data, the feed-forward neural network 'newff' has been used. The neural network is consist on one hidden layer and one output layer.

Throughout the training and testing of the neural network; various activation function (also called transfer function that defines the output of node on given input or set of inputs) like ‘tansig’, ‘logsig’ and ‘purelin’ has been used. While the 'traingdx', 'trainr', and ‘trainbr’ were used as training functions of the neural network.

After the training, the neural network was tested on various data proportions accuracy was analysed by comparing the actual result with targeted result. The analysis of relevant experiments is given below.

Experimental Results and Analysis

To understand the functionality of different parameters of the neural network, activation functions, training function and learning functions, various hypothesis are made to critically analyse the performance of the neural network. Then on the base of these hypothesis various experiment are made. In each experiment number of epochs are kept same which is 300.

Some hypothesis and experiments are presented below:

Role of Learning Rate in the Training

Hypothesis

The selection of learning rate matters a lot in the training of the neural network. It is supposed that for the better performance of the neural network in regard of training and accuracy, the learning rate should be small.

Experiment

- No. of Hidden Layer Neurons = 10
- No. of Hidden Layers = 1
- No of Output layer = 1
- Data Proportion = Training Data (70%), Testing Data (30%)
- Activation Functions = “tansig , tansig”
- Training Function = ‘trainbr’
- Learning Function = ‘learngd’
- Number of Epochs = 300

Learning Rate	Performance	Accuracy (%)
0.0001	2.18	93.7
0.001	1.01	96.5
0.01	1.92	94.01
1	1.54	94.7
2	0.68	95.8
5	2.16	95.5
10	1.57	95.00

Result

The results of the experiment has been shown above. Although there is no major difference in accuracy on changing learning rate (It may be due to activation function and training function). But still there is variations in the results. Which show that if learning rate is small the accuracy will be

maximum (with learning rate 0.001 in above experiment) and opposite case with high learning rate. The reason is that with large learning rate the weights proceed too far in right direction, which prevent network from better learning, then ultimately the accuracy will suppress.

So it can be concluded that small learning rate increases accuracy but too small learning rate can adversely affect the accuracy of network [11], which can also be seen in above experiment(With learning rate 0.0001 as compare to 0.01) .

Role of Activation Function

Hypothesis

Activation function plays a major rule in the training of the neural network. Every activation function should perform differently, because each has its own mathematical characteristics. Although, there is no much variation in the dataset, but based on mathematical composition of the activation functions, the accuracy should be reduced on using logsig at output layer.

Experiment

- No. of Hidden Layer Neurons = 10
- No. of Hidden Layers = 1
- Data Proportion = Training Data (70%), Testing Data (30%)
- Activation Functions(used in the experiments) = tansig, logsig, purelin
- Training Function = 'trainbr'
- Learning Function = 'learngd'

Hidden Layer	Output layer	Accuracy (%)
logsig	tansig	97.2
tansig	logsig	20.30
tansig	tansig	96.30
Purelin	Purelin	88.8
Purelin	Tansig	97.5
Purelin	Logsig	20.8
Logsig	Logsig	19.9

Result

We have seen various transfer functions on hidden and output layer. Our hypothesis is correct to great extent. The accuracy was fairly well when we keep on changing activation functions, but whenever we use the logsig function at output layer, there was tremendous reduction in the accuracy of the network.

The reason is that, logisg is non-linear transfer function, which returns the value in the range of 0 to 1. Therefore the outputs of the networks are limited to small range and network maps the output in the range of 0 and 1, which in result lose accuracy.

Role of Data Proportion

Hypothesis

If we increase the amount of training data, the accuracy will be improved. Because if we use more data in training, then our network will get training on a lot of possible cases, which will cause to improve accuracy of the network.

Experiment

- No. of Hidden Layer Neurons = 10
- No. of Hidden Layers = 1
- Data Proportion = Training Data (70%), Testing Data (30%)
- Activation Functions = 'logsig' , 'tansig'
- Training Function = 'trainbr'
- Learning Function = 'learngd'

Training Data	Testing Data	Accuracy (%)
50	50	97.4
70	30	97.1
90	10	97.7
10	90	66.2
40	60	95

Result

Hence it proved that if we use more data in training, the accuracy will be better, because our network will get training on many unseen cases, so ultimately the performance will high.

Effect of Number of Neurons in Hidden Layer

Hypothesis

If we use more neurons in the hidden layer, overall performance of the network will be increased.

Experiment

- No. of Hidden Layers = 1
- Data Proportion = Training Data (70%), Testing Data (30%)
- Activation Functions = 'logsig' , 'tansig'
- Training Function = 'trainbr'
- Learning Function = 'learngd'

No's of neurons on hidden layer	Performance	Accuracy (%)
1	0.4	95.1
3	2.04	96.2
5	0.95	95.8
10	2.2	96.5
15	1.20	94

20	2.6	96.00

Result

So it can be concluded from above experiment that if we increase the number of neurons, our performance will increase because in that case there will more neurons available for transferring data from hidden layer to output layer. But it can be seen from above experiment that more increase in number of neurons can supress the accuracy due to over fitting.

Conclusion

Artificial neural network is one of the best and widely used artificial intelligence based technique that can be used for diagnosing benignity and malignity of breast cancer. In that work various approaches are made to get best results using neural networks. It can be concluded that accuracy of the neural network depends on various factors: activation functions, training and learning algorithms and especially on distribution of dataset.

Bibliography

[1] Asieh Khosravanian and Saeed Ayat, "Diagnosing Breast Cancer Type by Using Probabilistic," *International Journal of Knowledge Engineering, Vol. 2, No. 1, March 2016.*

[2] Der-Ming Liou and Wei-Pin Chang, [Online]. Available: https://link.springer.com/protocol/10.1007/978-1-4939-1985-7_12.

[3] [Online]. Available: https://www.google.com.pk/search?q=Artificial+neuron&source=lnms&tbm=isch&sa=X&ved=0ahUKEwjjkv3F_OvXAhWL8RQKHRq-B0cQ_AUICigB&biw=1366&bih=588#imgrc=y--CpZwRgEVUHM:.

[4] [Online]. Available: https://www.google.com.pk/search?q=neural+network&source=lnms&tbm=isch&sa=X&ved=0ahUKEwjD0sWI--vXAhXOI1AKHS0vDXAQ_AUICigB&biw=1366&bih=637#imgdii=e0jyP8eElwTY9M:&imgrc=EXoHkcJRnWRI3M:.

[5] [Online]. Available: http://www.who.int/cancer/events/breast_cancer_month/en/.

[6] SHELLY GUPTA , DHARMINDER KUMAR ,ANAND SHARMA , "DATA MINING CLASSIFICATION TECHNIQUES APPLIED FOR BREAST CANCER DIAGNOSIS AND PROGNOSIS," *Shelly Gupta et al./ Indian Journal of Computer Science and Engineering (IJCSE).*

[7] [Online]. Available: https://docslide.net/documents/local-linear-wavelet-neural-network-based-breast-tumor-classification-using.html.

[8] D. W. H. Wolberg. [Online]. Available: http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Original).

[9] [Online]. Available: http://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/.

[10] [Online]. Available: http://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/breast-cancer-wisconsin.names.

[11] D. Randall Wilson and Tony R. Martinez, "The Need for Small Learning Rates," in *In Proceedings of the 2001 International Joint Conference on Neural Networks (IJCNN'01), 115-119.*, Utah, USA.

[12] [Online]. Available: http://www.who.int/cancer/events/breast_cancer_month/en/.