

Breast Cancer Classification Using

Neural Networks

A detailed report of classification of breast cancer data using neural networks

Sheharyar Saleem – 15026388

27th November, 2018­­

Contents

[Introduction: 3](#_Toc531555062)

[Background: 3](#_Toc531555063)

[Neural Networks: 3](#_Toc531555064)

[Input layer: 3](#_Toc531555065)

[Hidden layer: 4](#_Toc531555066)

[Output layer: 4](#_Toc531555067)

[Some more information: 4](#_Toc531555068)

[csvread: 4](#_Toc531555069)

[feedforward: 4](#_Toc531555070)

[setwb: 4](#_Toc531555071)

[Description of the Methodology: 4](#_Toc531555072)

# Introduction:

Cancer has become a very leading disease found in human race. If I describe cancer in one sentence I would say that the division of cells at abnormal rate is called cancer. There are many reasons a person can have cancer. Most common cancer found in males, is lung cancer which is caused by smoking cigarettes. And in females the most common type is breast cancer, which is caused by the same reason which is division and multiplying if cells in a chaotic manner.

After years of research, medical science has come up with ways to fight cancer, including breast cancer. The common ways to kill cancer cells are radiation, surgery, chemotherapy, targeted therapy etc. But to start any of the previous mentioned ways, we need to predict at which stage we are standing while operating the patient, so for that we need data. Hospitals and medical research centers have data, and using that data it becomes much more easy to start treatment of cancer.

# Background:

Breast Cancer is of two types, and to differentiate between them, we need data on which we could give our judgement. But before getting into that, we need to know difference between two cancers. One type is called Benign and the other is called Malignant. In Malignant, the cells divide which means this is cancerous type. So we need serious medical support. If we talk about Benign, in this case we do find tumor but this tumor does not divide, so we do not consider this cancer.

# Neural Networks:

If we talk about neural networks, neural networks were made by observing human brain. In human nervous system, nervous system detects a certain signal, and respond according to that signal, and with time nervous system learns each signal and in future acts accordingly. In computer, a neural network has neurons, organized in layers. This layers makes a neural network gain complexity. Layers can be of 3 types;

## Input layer:

This layer is made of artificial neurons and is responsible for getting data inside our system. The further work starts from here.

## Hidden layer:

This layer is used increase accuracy of our neural network. it takes in its probabilistic input signals, works on them and converts them into an output.

## Output layer:

This layer just outputs the result that is passed to it by hidden layer (if any used).

## Applications:

Neural networks have many applications. I am listing down a very few over here.

1)Speech recognition.

2)Character recognition.

3)Human face recognition.

# Some more information:

Before getting into practical work, I would like to clear some basic concepts. For the task given I would use matlab. To understand the stuff mentioned below one must know basics of matlab. Some important functions that I used in my project are as follows,

## csvread:

I have stored my data inside a csv file, to open the content of that file I used the function csvread.

## feedforward:

Feedforward is a neural network in which data is passed and output is received.

## setwb:

This function is used to set weights of the network.

## Feedforwardnet:

This is a deep learning model. To keep it simple, the information flows through the function which is being evaluated, and after computation it is given as output.

# 

# Description of the Methodology:

Following steps would be followed to solve the given problem.

1. First of all, we have to change our dataset a little. There are some fields missing in our data. There are question marks {?}. So first of all we will replace these question marks {?} with zeros {0}.
2. Load all the in a matrix, starting from column 2 to 10, and select rows from 1 to 350, we will train the network using this data.
3. In the dataset, first column has IDs of all the patients. For us this information is of no use. We will start picking data from the next column.
4. One more thing, last column has the result in side it, we will ignore that column too because we just need inputs. If you notice, there are two unique values, 2 and 4. 2 means benign and 4 means Malignant.
5. We will make an array of 2 columns in which only 0 and 1 would be stored, these 0s and 1s would differentiate between Malignant and Benign.
6. Once we train the network, we will calculate accuracy.

# Hypothesis:

So we have developed a very detailed foundation of the problem. Now we apply different techniques on our network to analyze its performance.

## Learning rate:

Learning rate controls how much we are adjusting our network with respect to loss gradient.

## Hypothesis:

The user generally sets the value of learning rate. It is observed that if the value is kept medium the accuracy increases.

## Results:

|  |  |
| --- | --- |
| Learning rate | Accuracy |
| 0.5 | 95.98 |
| 0.01 | 96.78 |
| 0.001 | 23.78 |
| 1.0 | 23.69 |

## Conclusion:

It’s simple that if we keep the learning rate medium, accuracy increases. If learning rate is kept low, then training becomes more reliable, but optimization will take a lot of time. If learning rate is kept high the training might separate.

## Hidden Layers:

I have already discussed what hidden layers are. We can increase hidden layers, this might help us, but we are not sure, because this mechanism also depends on complexity of the problem. One more problem that we can face is that if we increase hidden layers more than a specific number the accuracy might drop down. It will be able to learn the data but when new data would arrive it won’t work.

## Results:

|  |  |
| --- | --- |
| layers | Results |
| 10,10 | 96 |
| 10,10,10 | 96 |
| 100,100,100 | 95 |
| 1000,1000 | 23 |

## Conclusion:

Accuracy decreases if we increase the number of hidden layers. And the accuracy depends on the size of data we have, and according to that data we will add hidden layers