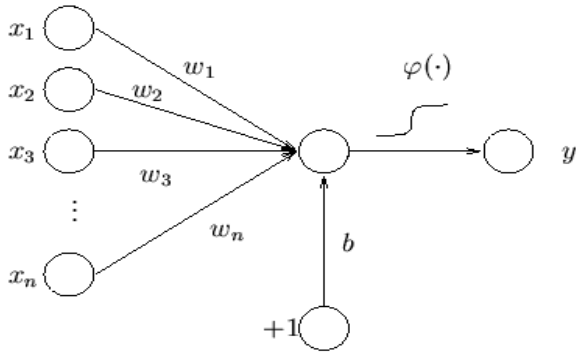
Multi Layer Perceptron on Breast Cancer Wisconsin dataset

Shashank Pathak(15BCE1287)*,
*School of Computing Sciences and Engineering, VIT Chennai, Tamilnadu, India 600127
Email: shashank.pathak2015@vit.ac.in*

***Abstract*—In machine learning, an MLP is a network of simple neurons called perceptrons. The perceptron computes a single output from multiple real-valued inputs by forming a linear combination according to its input weights and then possibly putting the output through some nonlinear activation function.We study the performance of the MLP for the classification of the Breast Cancer Wisconsin Dataset.The system developed performs with 96.65% insample accuracy and 97.56% accuracy for the test sample in a two class classification**



1. Introduction

A multilayer perceptron (MLP) is a class of feedforward artificial neural network. An MLP consists of at least three layers of nodes. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. MLP utilizes a supervised learning technique called backpropagation for training.Its multiple layers and non-linear activation distinguish MLP from a linear perceptron. It can distinguish data that is not linearly separable.

2. Methodology

A single perceptron is not very useful because of its limited mapping ability. No matter what activation function is used, the perceptron is only able to represent an oriented ridge-like function. The perceptrons can, however, be used as building blocks of a larger, much more practical structure. A typical multilayer perceptron (MLP) network consists of a set of source nodes forming the input layer, one or more hidden layers of computation nodes, and an output layer of nodes. The input signal propagates through the network layer-by-layer.This forms a multi layered perceptron structure.

3. Database - Breast Cancer Wisconsin Dataset(Original)

The Wisconsin Breast Cancer datasets from the UCI Machine Learning Repository is used,to distinguish malignant (cancerous) from benign (non-cancerous) samples.This dataset consist of 699 instances and 11 attributes that would help the classiy the data into the two classes.The attributes are described in the fig.1

#	Attribute	Domain
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)

Figure 1. Breast Cancer Wisconsin Dataset Attributes

4. Algorithm

An MLP is a network of simple neurons called perceptrons.The perceptron computes a single output from multiple real-valued inputs by forming a linear combination according to its input weights and then possibly putting the output

through some nonlinear activation function. Mathematically this can be written as

$$y = \varphi(\sum_{i=1}^{n} w_i x_i + b) = \varphi(\mathbf{w}^T \mathbf{x} + b)$$

where $\mathbf{w}$ denotes the vector of weights, $\mathbf{x}$ is the vector of inputs, b is the bias and φ is the activation function

.

5. Experiments

First we used NeuralNetMLP from scikit learn library to classify the dataset on a reduced set of attributes namely:

'ClumpThickness', 'UCellSize', 'UCell-Shape','MargAdhes', 'EpiCellSize', 'BareNuc', 'Chromatin', 'Nucleoli' and 'Mitoses'. After this I did some preprocessing i.e. scaling the data and taking the standardized values.After this I trained the Neural Network for 1000 epochs and Then I plotted the graps between costs and epochs using matplotlib

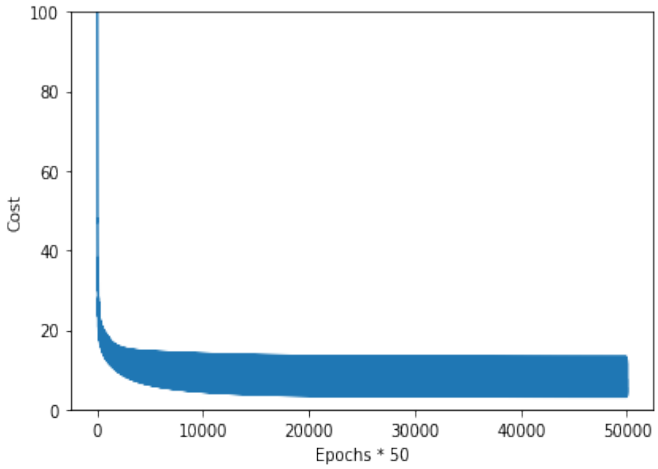


Figure 2.

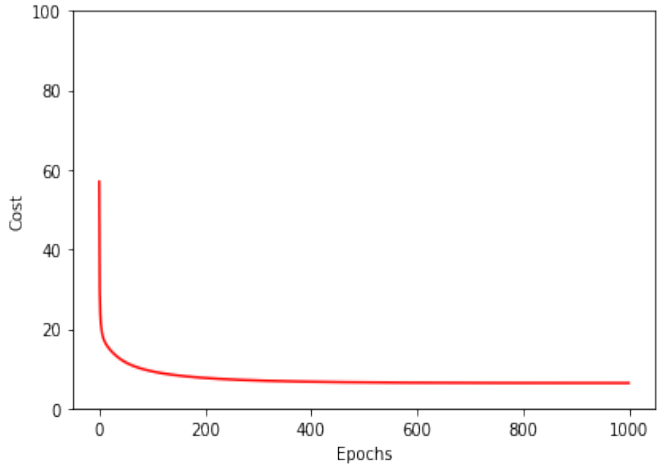


Figure 3.

6. Conclusion

After the classification we can see that the classification was very effective,.This indeed suggest that the MLP model is ver apt to classifify the data into two class Benign and Malignant for the breast cancer dataset The system developed performs with 96.65% insample accuracy and 97.56% accuracy for the test sample in a two class classification