CSE 574 Machine Learning
Programming Assignment 1
Handwritten Digits Classification

Due Date: March 8th 2017

Group No 67

Aniruddh Chaturvedi [5020 6958]
Arnav Ahire [5020 8006]
Ashwin Nikam [5020 7368]

1. Choosing hyper-parameter for Neural Network:

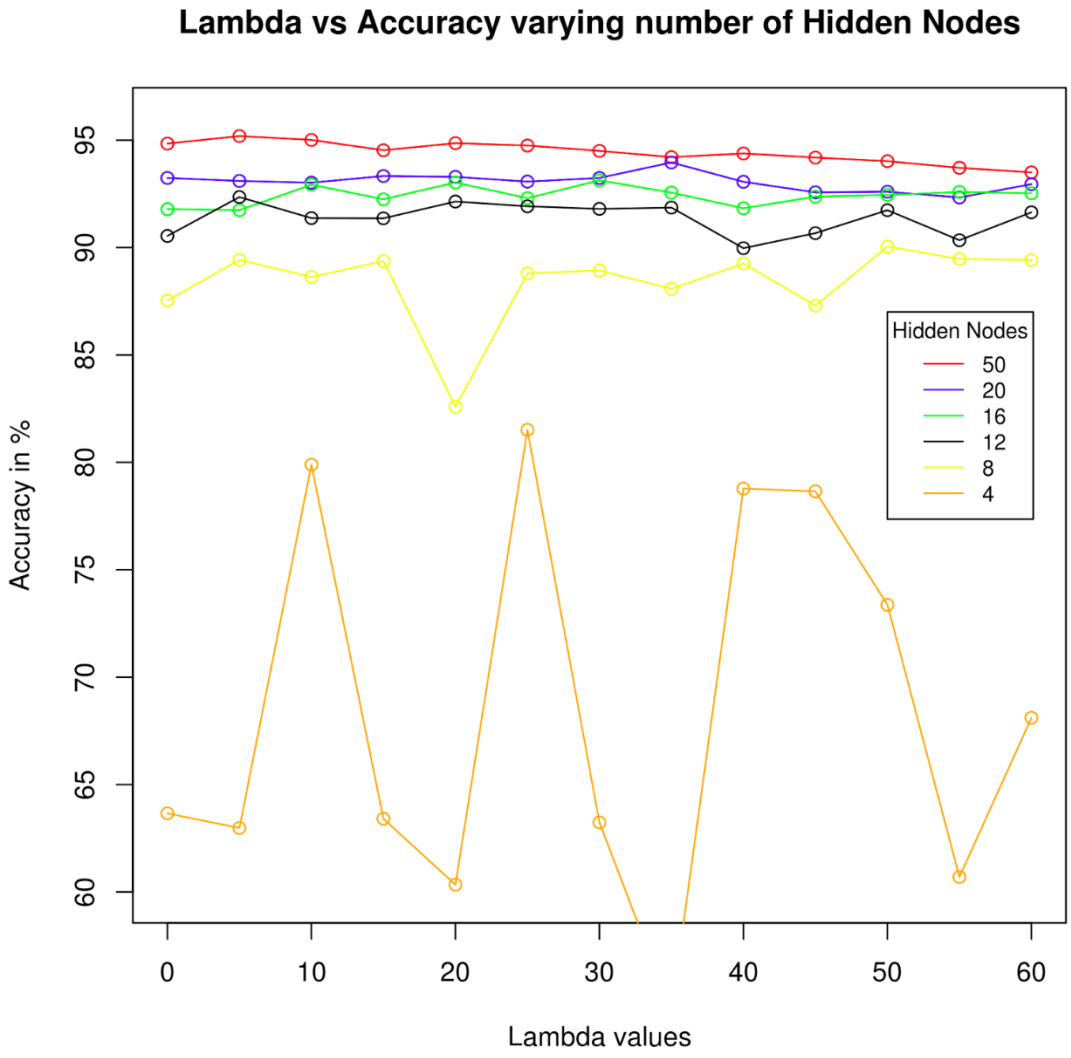
The neural network performance depends upon the following parameters:

- Number of Hidden Nodes
- Regularization Term- Lambda

In order to choose the hyper-parameters for optimum performance, we varied number of Hidden Nodes and value of regularization term i.e Lambda.

We varied the number of Hidden Nodes as follows:
Number of nodes: 4,8,12,16,20,50

We varied regularization term as follows:
Lambda: 0,5,10,15,20,25,30,35,40,45,50,55,60
An overview graph of Lambda versus Accuracy with varying number of Hidden Nodes is as follows:



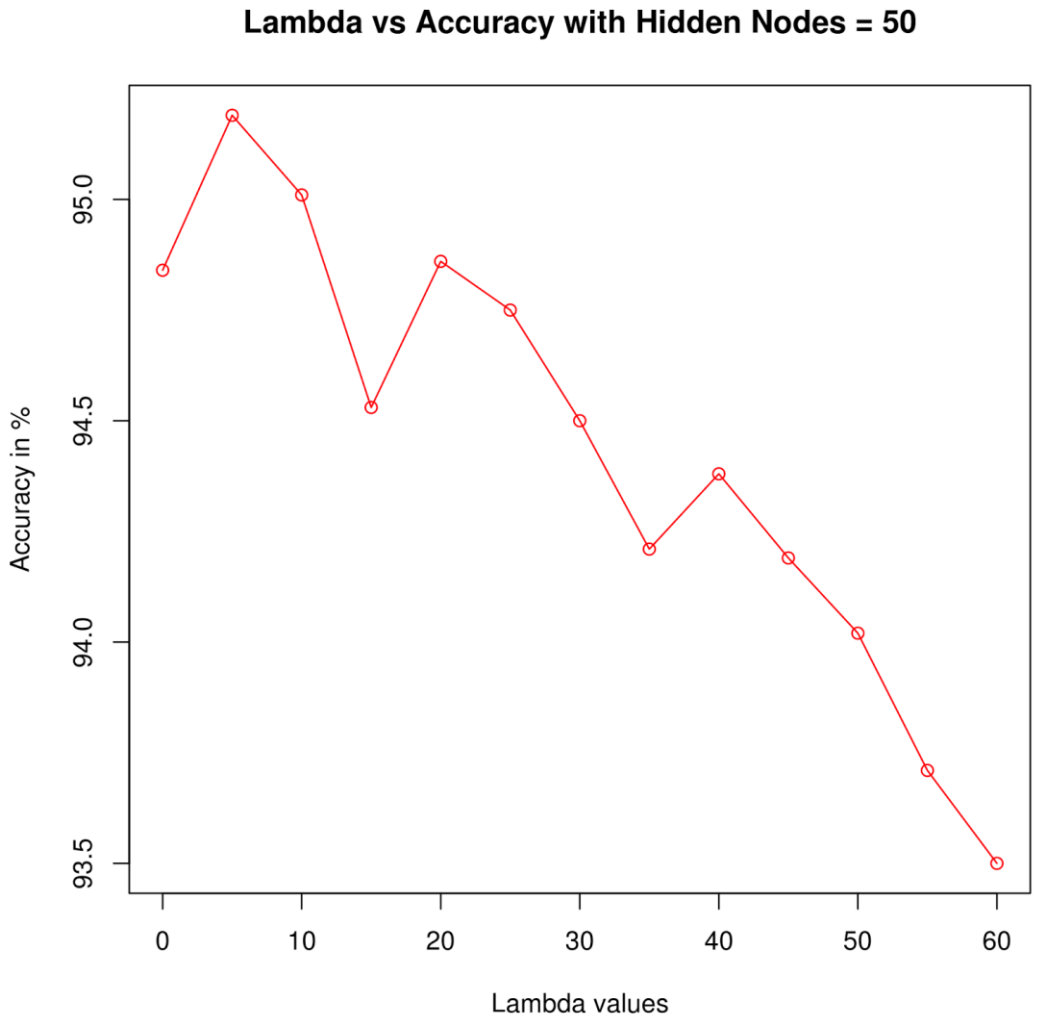
Observations:

1. This graph gives an overview of performance of neural network with varying parameters that is the number of Hidden Nodes and value of Regularization term - Lambda.
2. The accuracy of the neural network is inconsistent for hidden nodes that are less than 12. The graph however, stabilizes as the number of hidden nodes increase. This implies that more the number of hidden nodes in the hidden layer, greater is the obtained accuracy.
3. The accuracy of neural network is maximum for Hidden Nodes=50 and for Lambda=5.

Let us see how we obtained these values.

1. Plot for Hidden Nodes=50

   We drew a plot by keeping the number of hidden nodes constant (in this case 50). We checked for the greatest value of lambda for which we obtain the maximum accuracy. The following is the resultant graph.

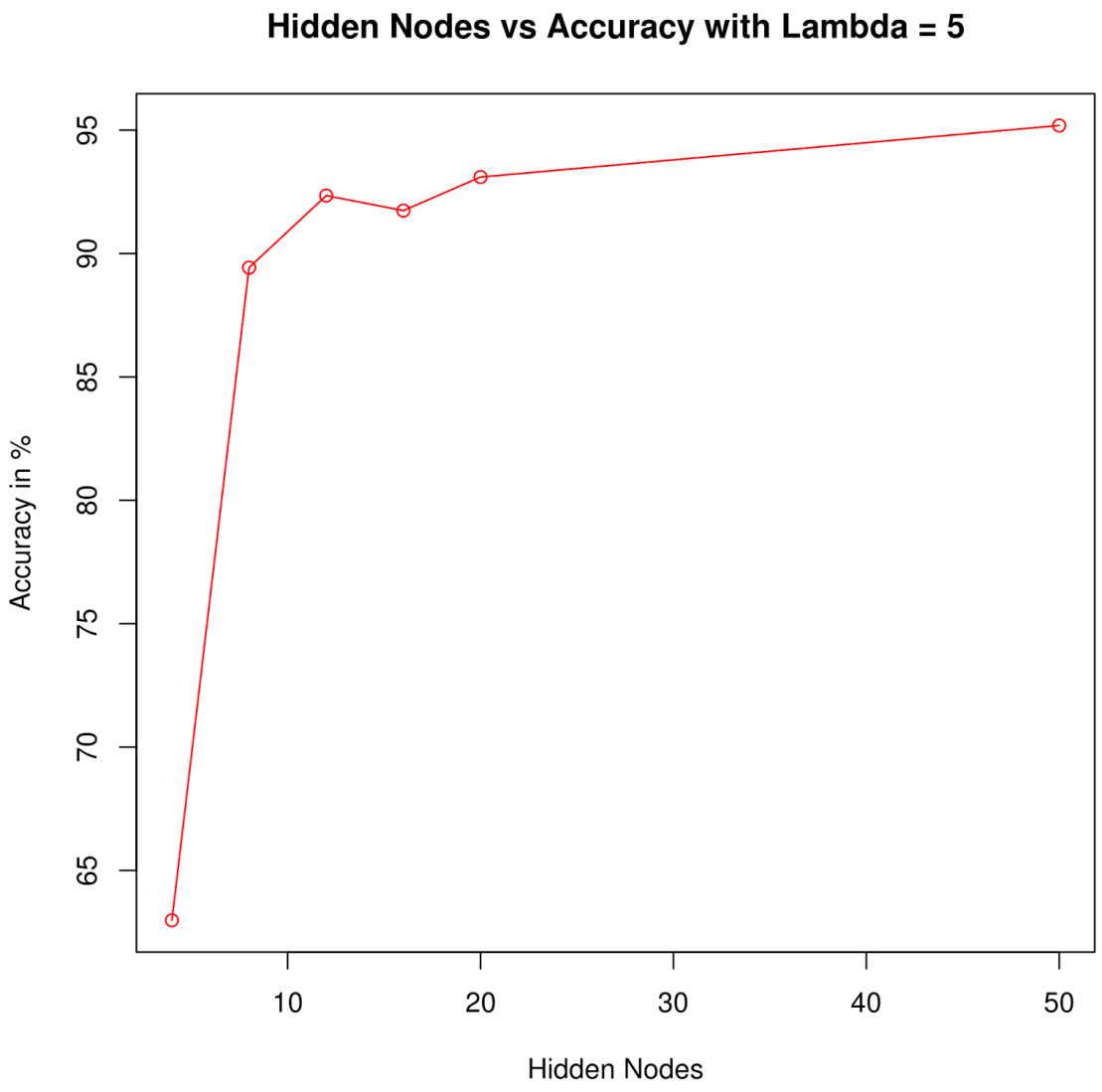


Observations:

1. We observe from the above graph that as Lambda value increases, the accuracy increases up to a threshold value after which any increase in Lambda value decreases the accuracy of the neural network.
2. We obtain the maximum accuracy for neural network when Lambda=5, after which any increase in the Lambda value is causing the accuracy to decrease. Thus we obtain the hyper-parameter for Lambda as 5.

2. Plot for Lambda=5

We plot another graph where we kept Lambda=5 as constant and changed the number of Hidden Nodes. The graph obtained is as follows:



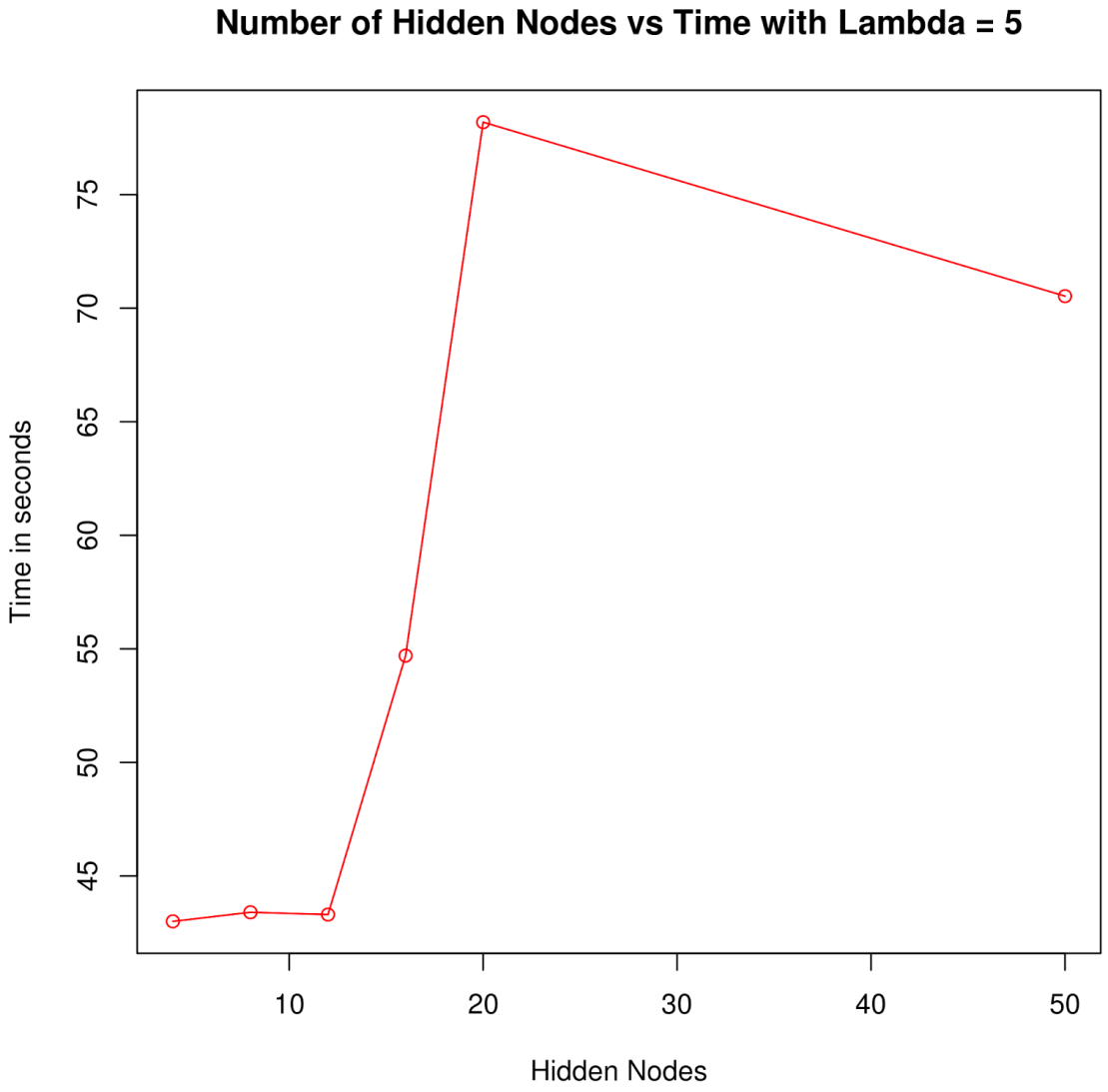
Observations:

1. From the above graph we can clearly observe that as the number of hidden nodes increase in the hidden layer, the accuracy in the neural network increases as stated earlier.

2. This also verified our selection of hyper-parameter i.e number of hidden nodes as 50 since we obtained the maximum accuracy for Hidden Nodes=50 in the above graph.
3. Hence we have our hyper-parameters as follows:
    - Hidden Nodes=50
    - Lambda (Regularization term) =5

2: Accuracy of classification method on handwritten digits' test data

We obtained the accuracy against the training time as follows:



Observations:

1. Keeping Lambda=5 as constant, we checked what happens to the training time of the neural network when the number of hidden nodes are increased.
2. As we can see from the graph, the training time for the neural network increases as we increase the number of Hidden Nodes in the neural network.
3. In order to reduce the time to train the data, we need to reduce the number of hidden nodes, however doing so would reduce the accuracy of the neural network too so reducing the number of hidden nodes to a very small number should also be avoided.

3: Comparison of Neural Network with Deep Neural Network

We computed the accuracy of our neural network for CelebA data set (similar to what we did for MNIST DataSet). Since our neural network consists of only 1 hidden layer, we obtain the accuracy for the set of obtained hyper-parameters.

In this section we have also computed the accuracy of Deep Neural Networks using the TensorFlow library for multiple hidden layers. Accuracies have been computed for 3, 5 and 7 hidden layers along with the default 2 hidden layers.

Computations were performed on the Springsteen server which already had Python 3.4.3 and TensorFlow 0.12.1 installed. The computations were carried out on the CelebA data set which was used previously on the single layer neural network.

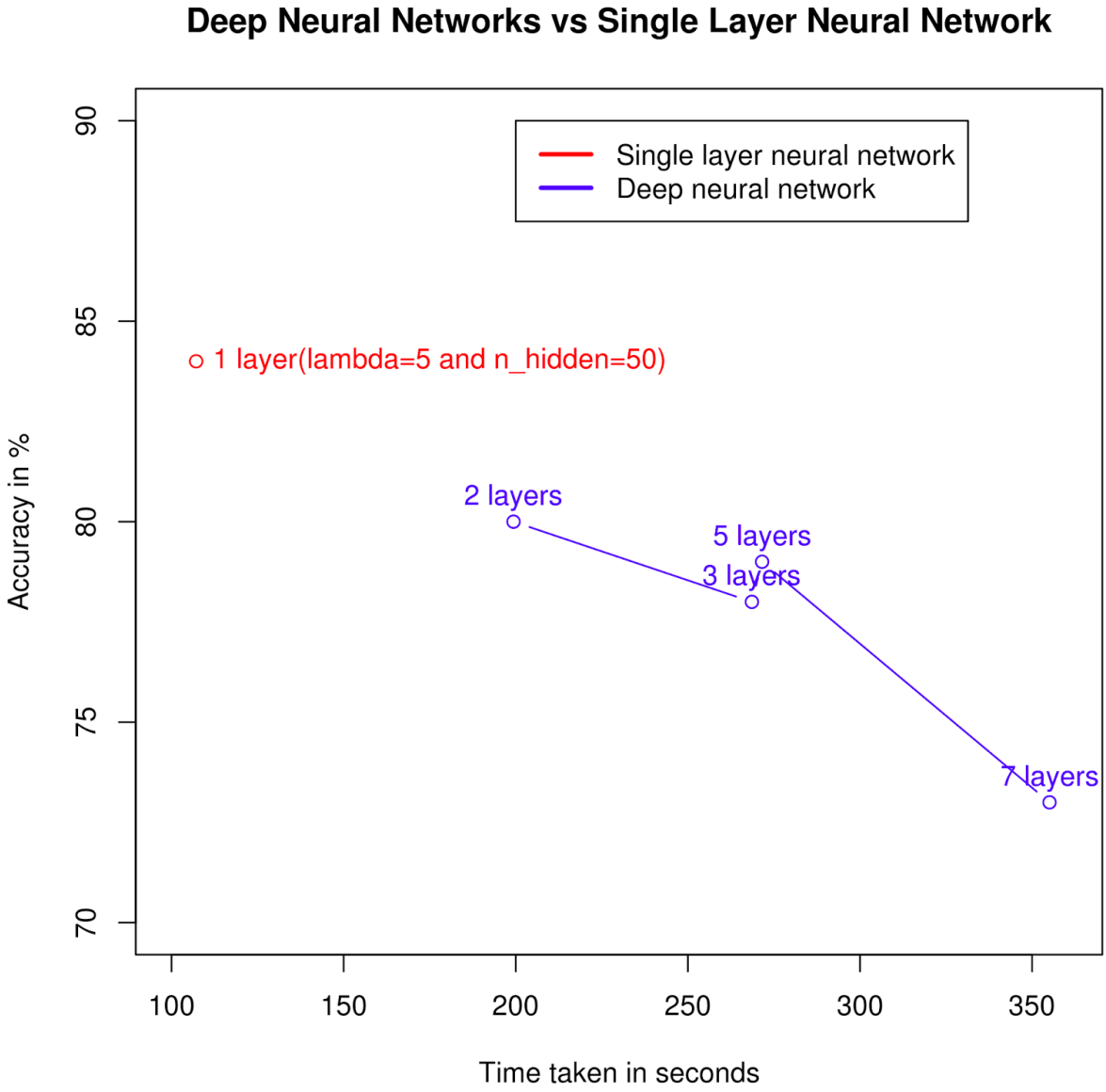
Results show that as we increase the hidden layers, the time for computation also increases. However, the accuracy not necessarily increases as we increase the hidden layers.

This is because of the principle of Occam's Razor as stated by the professor which says that 'More complicated networks will not give necessarily better results.' The comparison between a single layer neural network and deep neural network visualizes this as shown in the graph below.

The lambda value for the single layer neural network has been set to 5 and the number of hidden nodes have been set to 50 which are the hyper-parameters.

From the given graph, we can observe that increasing the layers in deep neural network is directly proportional to time. However, the accuracy keeps on decreasing. (Except for a slight bump from 3 to 5)

By carefully observing the graph we can conclude that single layer neural network is showing a much better accuracy than all the deep neural networks and also is getting computed in a much better time.


Deep Neural Networks vs Single Layer Neural Network
Single layer neural network
Deep neural network
1 layer(lambda=5 and n_hidden=50)
2 layers
3 layers
5 layers
7 layers
Accuracy in %
Time taken in seconds
90
85
80
75
70
100
150
200
250
300
350