CSE 574 Machine Learning
Programming Assignment 2

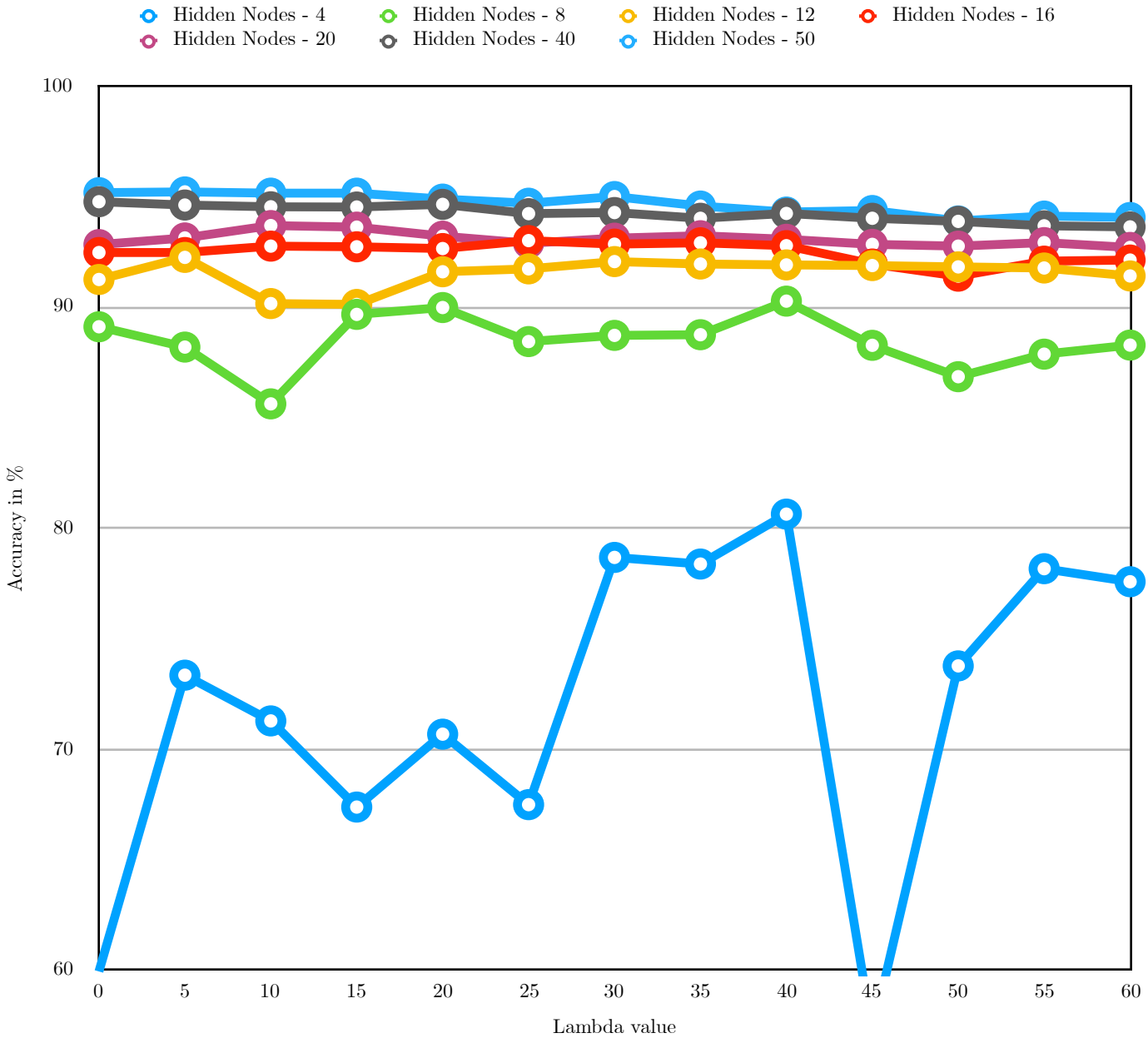
Yesh Kumar Singh - 50247208
Sneha Mehta - 50245877

1. The hyper parameters of the neural network are as follows:
    1. The number of hidden nodes
    2. The value of lambda (Regularization term)

We varied the parameters as follows:

1. Number of hidden nodes: 4, 8, 12, 16, 20, 40, 50
2. Value of lambda: 0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60

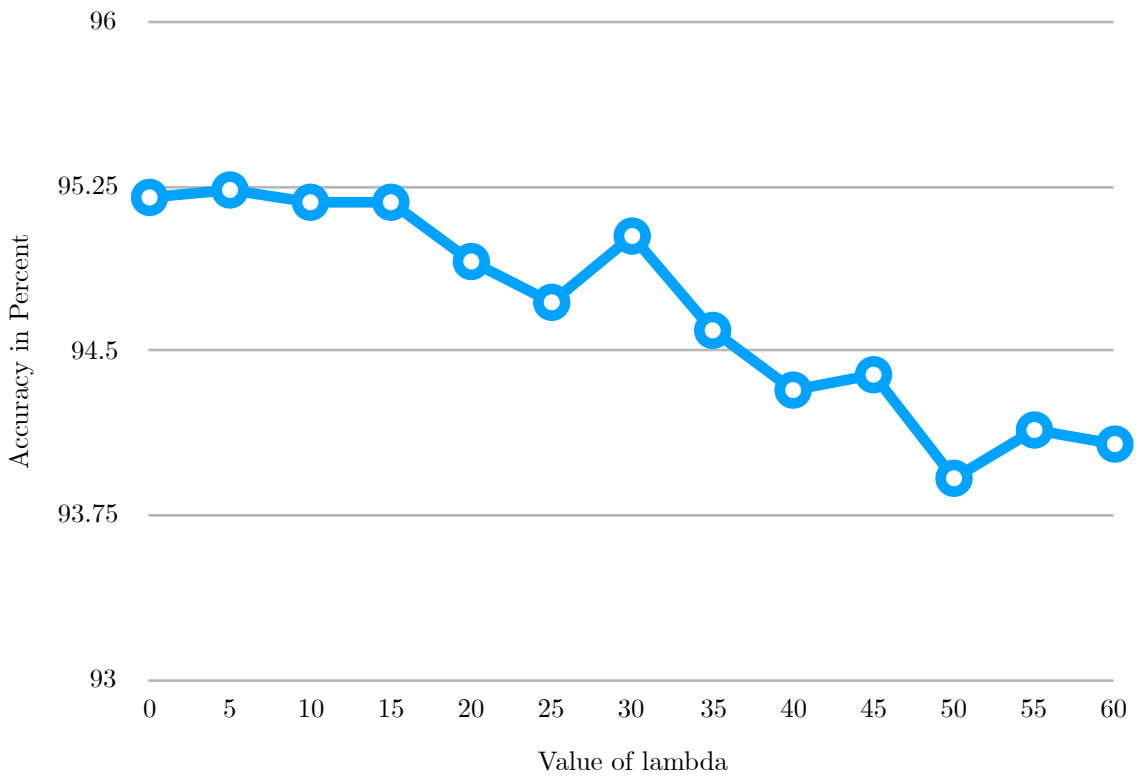
In order to compare the performance we plotted a graph of Accuracy on training data vs. Value of lambda for all types of networks i.e. with 4 hidden nodes to with 50 hidden nodes.



Conclusion:

1. The graph gives an insight in the performance of the neural network with varying parameters i.e. the number of hidden nodes and the value of lambda.
2. The performance in terms of accuracy of the neural network below 12 hidden nodes is very inconsistent to comment on.
3. The performance is the best for the neural network with 50 hidden nodes and the value of lambda as 5. Accuracy = 94.89 % for this value of lambda.

I. When hidden nodes = 50

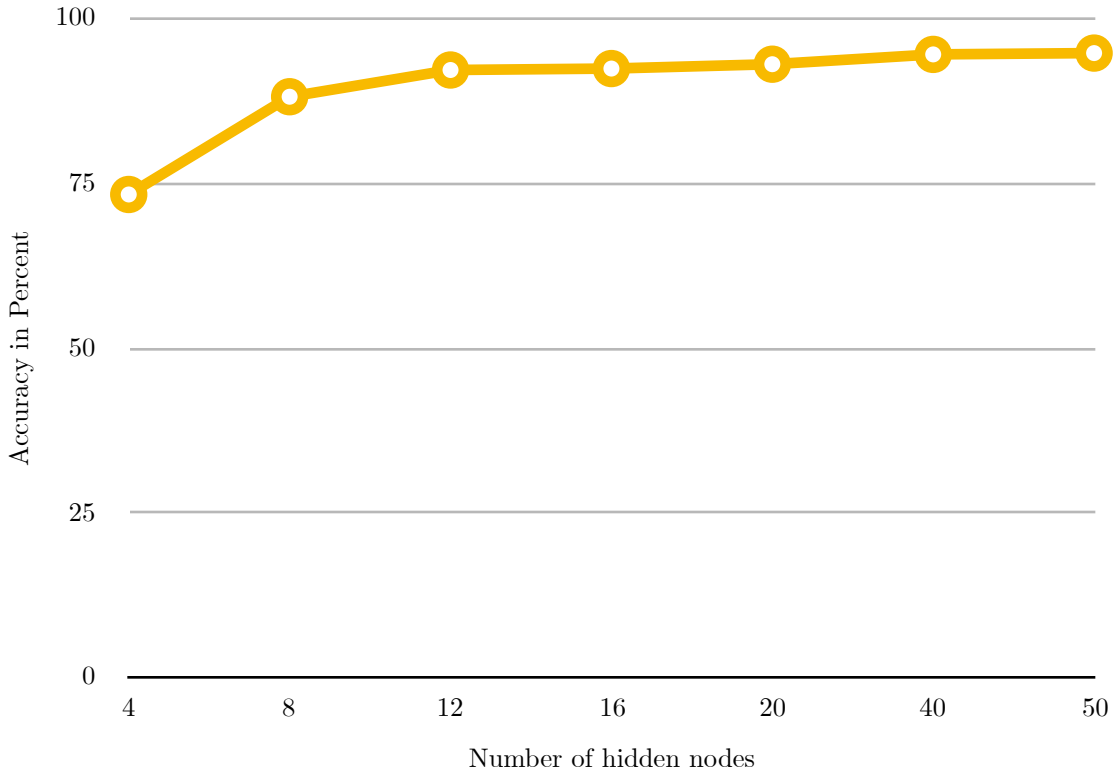


We can see that the accuracy on training data is maximum when value of lambda is 5 and as the value of lambda increases the accuracy starts to decrease gradually.

Conclusion:

1. For λ equal to 5 the neural network with 50 nodes perform the best.
2. As the value of λ increases the accuracy decreases with a few outliers ignored.

II. When the lambda is 5



At the particular value of lambda the values of the accuracy in percent is maximum in most of the cases i.e. with 40 hidden nodes, 12 hidden nodes etc.

Conclusion:

1. The value of $\lambda = 5$ is the optimal.

Final Conclusion:

- Our hyper parameters are $\lambda = 5$ and number of hidden nodes $= 50$

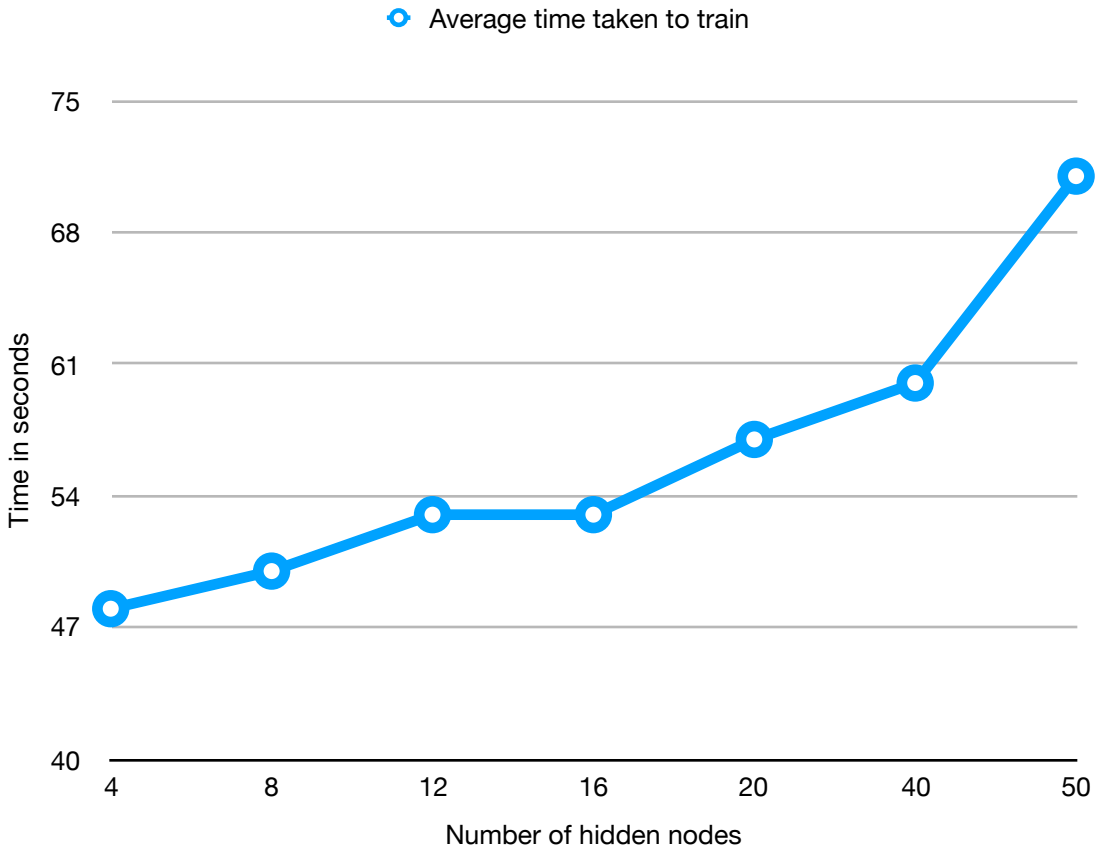
2. Accuracy of Classification method on the handwritten digits test dataset:

We fixed the number of hidden nodes = 50
And the value of $\lambda = 5$

For this we get the accuracies and timings as follows:

Time required to train the network	=	69.23412 seconds
Accuracy on the training set	=	95.234 %
Accuracy on the validation set	=	94.18 %
Accuracy on the testing set	=	94.84 %

Average time taken vs Number of Nodes:



We can see that as the number of nodes in the hidden layer increases the time to train the entire neural network also kind of linearly increases.

3. Accuracy of the classification method on the CelebA data set:

The number of hidden nodes = 256

And the value of $\lambda = 10$
For this we get the accuracies and timings as follows:

Time required to train the network	=	81.23 seconds
Accuracy on the training set	=	84.587 %
Accuracy on the validation set	=	83.302 %
Accuracy on the testing set	=	85.124 %

And the value of $\lambda = 5$
For this we get the accuracies and timings as follows:

Time required to train the network	=	87.32 seconds
Accuracy on the training set	=	84.905 %
Accuracy on the validation set	=	83.302 %
Accuracy on the testing set	=	84.935 %

And the value of $\lambda = 15$
For this we get the accuracies and timings as follows:

Time required to train the network	=	102.77 seconds
Accuracy on the training set	=	84.748 %
Accuracy on the validation set	=	83.039 %
Accuracy on the testing set	=	84.859 %

As we can see our accuracy is not going above 85.124 % even for two different values of lambda.

4. Comparison of our single layer Neural Network to DeepNN

We ran the following configurations of the network and the accuracy and the time taken by them were as follows:

Deep Neural Network with 2 layers:	
Accuracy	0.816427
Time Taken (in seconds)	121.42735600471497
Number of Epochs	100

Deep Neural Network with 2 layers:	
Accuracy	0.805829
Time Taken (in seconds)	66.99964785575867
Number of Epochs	50

Deep Neural Network with 3 layers:	
Accuracy	0.77025
Time Taken (in seconds)	67.69079613685608
Number of Epochs	50

Deep Neural Network with 3 layers:	
Accuracy	0.801665
Time Taken (in seconds)	150.07162499427795
Number of Epochs	100

Deep Neural Network with 5 layers:	
Accuracy	0.758895
Time Taken (in seconds)	155.95512866973877
Number of Epochs	100

Deep Neural Network with 5 layers:	
Accuracy	0.74754
Time Taken (in seconds)	89.71809101104736
Number of Epochs	50

Deep Neural Network with 7 layers:	
Accuracy	0.748297
Time Taken (in seconds)	198.08757972717285
Number of Epochs	100

Deep Neural Network with 7 layers:	
Accuracy	0.72483
Time Taken (in seconds)	89.92814087867737
Number of Epochs	50

The maximum accuracy that we obtained through the Deep NN was around 81.64 percent which is less as compared to the accuracy of the single layer network 85.12 %.

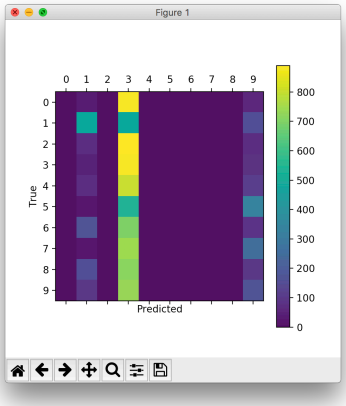
Comparison table:

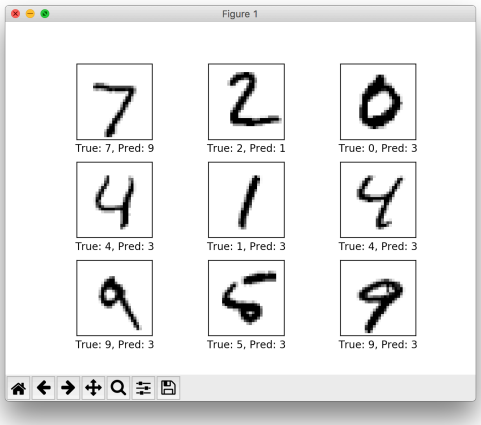
Features	Single layer Neural Network	Deep Neural Network
Number of layers	1	2
Accuracy	85.124%	81.6427%
Time taken (seconds)	81.23	121.42

Conclusion: As we can see from the table the accuracy decreases and the training time increases in the neural network. This is because of the principle of Occam's razor. A more complicated network doesn't mean a better network.

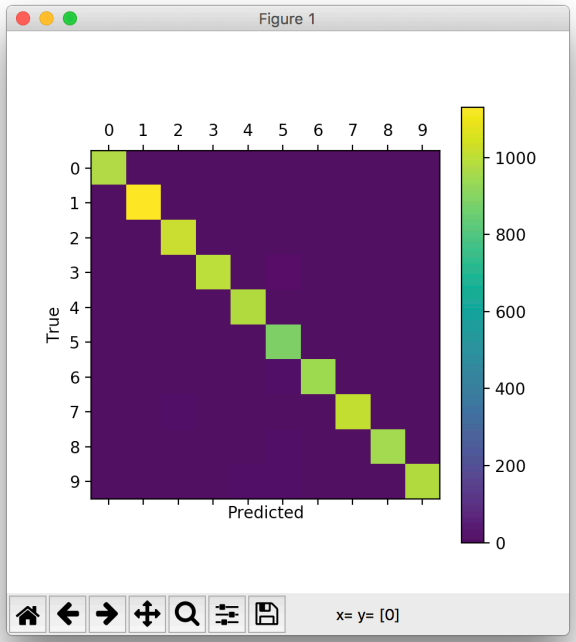
4. Results of the convolutional neural network
The data set is divided as follows:
1. Training set: 55000
2. Validation set: 5000
3. Test set: 10000

The first few iterations perform very poorly as expected and example of error and confusion matrices are as follows:





The final confusion matrix after 9901 iterations which takes 6 mins 4 seconds to complete and gives an accuracy of 98.8 % looks as follows:



The Confusion matrix is:

```
[[ 973    0    1    1    0    2    1    0    2    0]
 [   0 1131    1    0    0    0    2    1    0    0]
 [   2    1 1021    1    1    0    0    3    3    0]
 [   1    0    0  996    0    9    0    2    2    0]
 [   0    0    1    0  978    0    1    0    0    2]
 [   1    0    0    0    0  890    1    0    0    0]
 [   3    2    0    0    1    7  944    0    1    0]
 [   0    1    5    2    0    0    0 1015    1    4]
 [   3    0    2    2    1    7    0    2  955    2]
 [   3    3    0    1    8    7    0    3    3  981]]
```

The convolutional neural network takes a lot of time but provides a high amount of accuracy.