CSE 574

INTRODUCTION TO MACHINE LEARNING

PROGRAMMING ASSIGNMENT 1

Handwritten Digits Classification

By Group 38

Tulika Sengupta (tulikase@buffalo.edu)

Darshan Prakash Mhatre(dmhatre@buffalo.edu)

Divyansh Bhatnagar(dbhatnag@buffalo.edu)

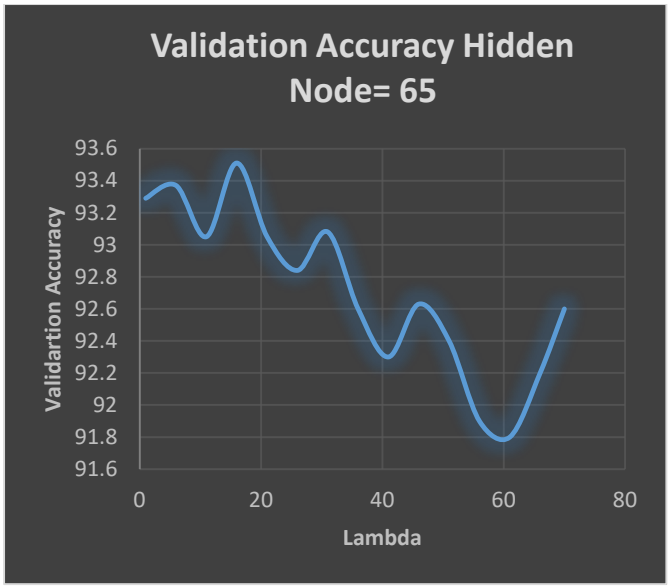
Objective: To implement a Multilayer Perceptron Neural Network and evaluate its performance in classifying handwritten digits.

Observations:

Regularization: The value of the hyper-parameter 'Lambda' regulates the regularization in neural Networks. It controls the bias variance to overcome the problem of overfitting. If we take lambda to be too small, then it means we are only considering in minimizing the error in the training data. So to control the bias variance we have to keep the value of lambda not too high nor too low.

Lambda	*Training Time*	*Validation Accuracy*	*Test Accuracy*
1	217.5	92.76	93.29
6	213.8	92.76	93.37
11	229.5	92.38	93.05
16	249.8	93.02	93.51
21	245.3	92.5	93.05
26	241.4	92.29	92.84
31	247	92.62	93.08
36	260.8	92.14	92.6
41	270	91.8	92.3
46	274.4	92.2	92.63
51	257.3	91.8	92.4
56	265.7	91.3	91.9
61	269.8	91.6	91.8
66	263.2	91.6	92.2
71	257.8	92.1	92.6

Table 1: Lambda Value for nnScript



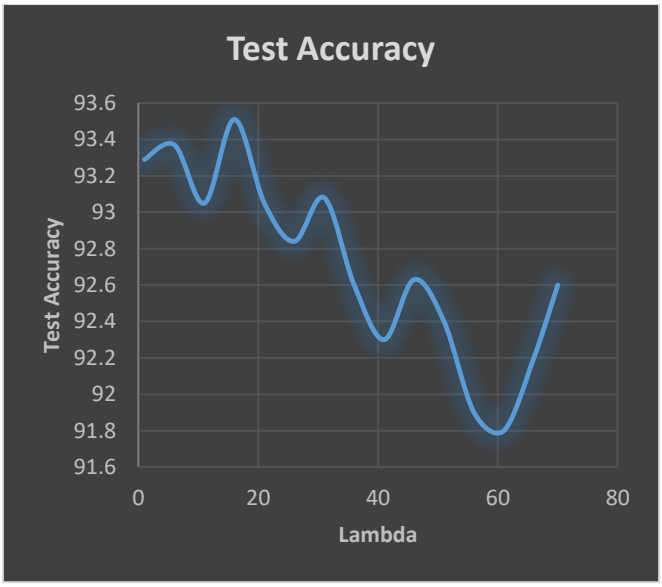


Figure 1: Validation Accuracy against Lambda

Figure 2: Test Accuracy against Lambda

Optimal Lambda Value = 16

The best number of hidden units depends in a complex way on:

- the numbers of input and output units
- the number of training cases
- the amount of noise in the targets
- the training algorithm
- regularization

If you have too few hidden units, you will get high training error and high generalization error due to under-fitting and high statistical bias. If you have too many hidden units, you may get low training error but still have high generalization error due to overfitting and high variance.

We tried it with different number of neurons in the hidden layer. **Starting with 50 and going till 70 for every cycle**, trained our network and tested it so we can capture the error rate. saved these data and decided how many hidden layer neurons do the best for us.

Hidden Nodes	*Training Time*
50	144
65	249.89
70	279.33
75	284
85	298.5

Table 2: Training Time for different Hidden Nodes

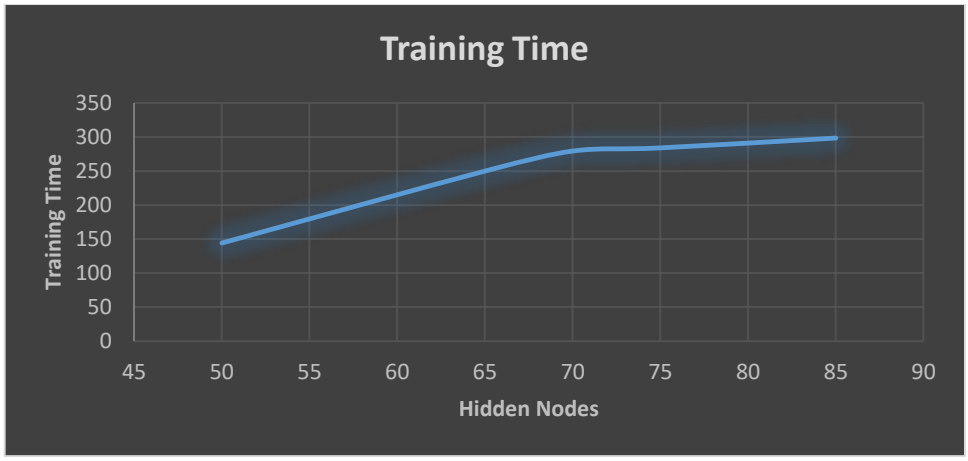


Figure 3: Training Time against number of Hidden Nodes

Result for MNIST dataset (nnScript):

The **accuracy for validation data set is 93.51**. We varied the number of hidden nodes from 50 to 85 and varied the regularization parameter from 0 to 70. We got the **optimal value for lambda = 16 and for number of hidden nodes = 65.**

Result for CelebA dataset (Facenn):

The **accuracy for validation data set is 81.** We varied the number of hidden nodes from 50 to 85 and varied the regularization parameter from 0 to 70. We got the **optimal value for lambda = 41 and for number of hidden nodes = 256.**

Comparision between the performance of single hidden layer neural network vs deep neural network:

Similar to nnScript, we tried different combinations of lambda and hidden nodes on CelebA dataset to distinguish between two classes- wearing glasses and not wearing glasses.

We got the **optimal value for lambda = 41** and for **number of hidden nodes = 256**

For hidden nodes= 256,

lambda	training time	test accuracy
1	82.65	67.34
6	185.47	78.82
11	195.95	71.42
16	250.3	70.28
21	205.94	67.63
26	254.39	74.82
31	359.93	77.56
36	743.75	70.96
41	579.51	81.68
46	547.93	78.61
51	297.55	74.11
56	416.28	80.24
61	355.86	76.87
66	322.53	74.64
71	347.38	72.89

Table 3: Lambda value for facenn

The above observations were for single neural network. To observe the performance under multiple hidden layer network, we used a TensorFlow library.

For different number of hidden layers, we found the following observations.

number of hidden layers	accuracy	training time
2	79.89	198.09
3	78.51	209.36
5	77.43	243.46
7	76.83	290.15

Table 4: Hidden Layer effect on accuracy and training time

Comparing these results with our optimised single neural network **which gave accuracy to 81.68**, we can conclude that though training time increases with increase in number of hidden layers, accuracy of neural network is not directly proportional to the number of hidden layers in the neural network. In fact, **accuracy is more dependent on the implementation of the hyper parameters like regularization co efficient and hidden nodes.**