CSE 574 Introduction to Machine Learning

Programming Assignment 1

Handwritten Digits Classification Using Neural Networks

Group 11 :

Ahut Gupta (50169273)

Rachna Shivangi (50169516)

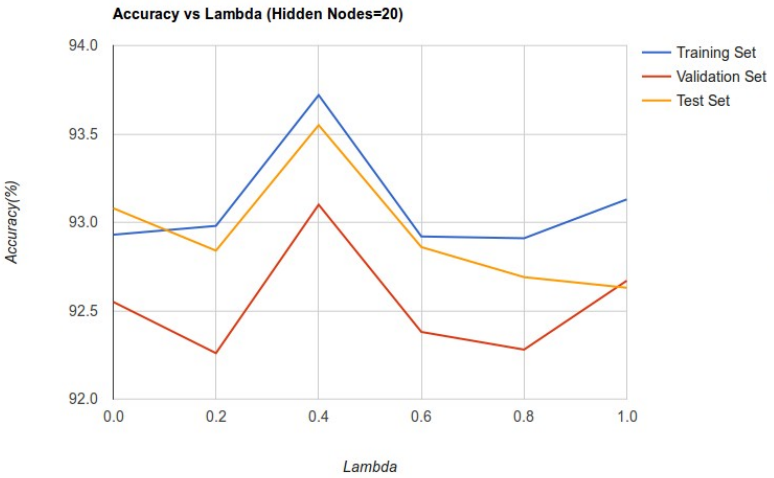
Vandana Chokkam (50167945)

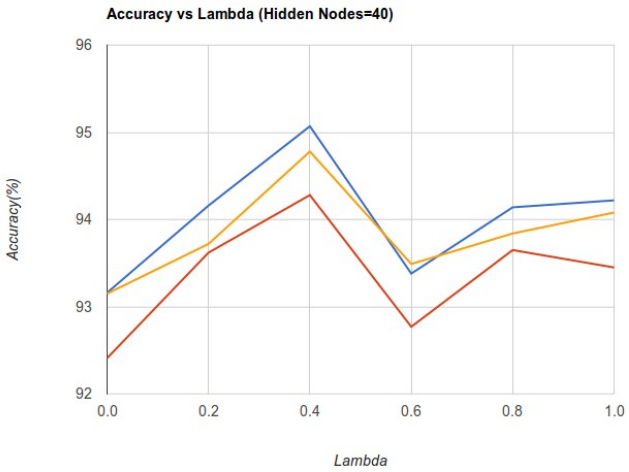
Abstract

In this assignment, we have implemented a Multilayer Perceptron Neural Network to classify handwritten digits from 0 to 9. In the preprocess function, feature selection effectively removes 67 of the 784 features from the training data set. The error function was regularized by a factor of λ.

Lambda vs. Accuracy

Iterating over the dataset for different number of hidden nodes (m): 4,8,10,16,20,40 and λ (0-1 incremented by 0.2), we observed the accuracy on training, validation and test sets. The highest accuracy was obtained when the hidden nodes were 20 and 40 for which the following graphs were plotted varying the values of lambda





From the first graph (m=20), we observe the problem of overfitting for λ <= 0.2 and λ > 0.6, wherein the accuracy for the training set increases, while it decreases for the validation and test sets.
For λ > 0.4, we observe underfitting, as the training set accuracy decreases.
Also, from the second graph (m=40) we observe that the highest accuracy (>94%) is achieved at λ = 0.4 and m = 40, with no overfitting.

The table lists the accuracies for all the datasets when m=40.

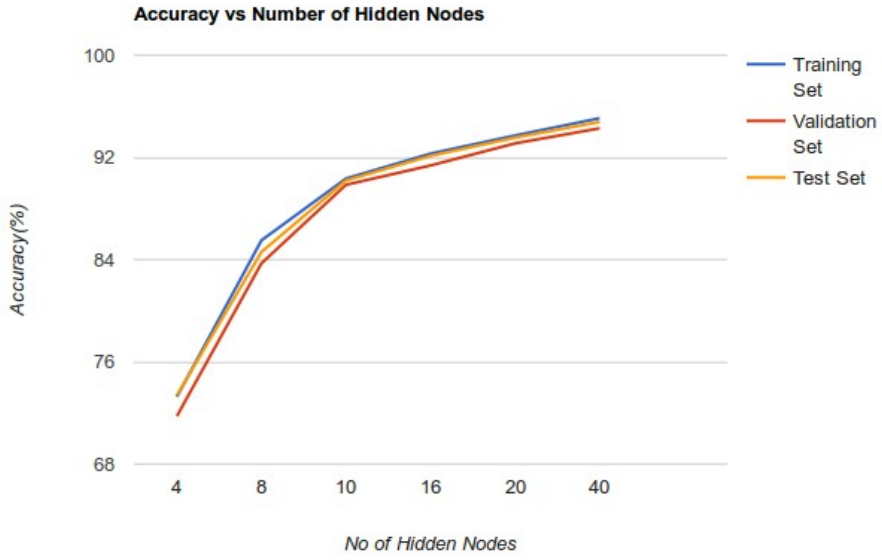
Lambda	Training Set Accuracy (%)	Validation Set Accuracy (%)	Test Set Accuracy (%)
0	93.16	92.41	93.15
0.2	94.16	93.62	93.72
0.4	95.07	94.28	94.78
0.6	93.38	92.77	93.49
0.8	94.142	93.65	93.84
1	94.22	93.45	94.08

Thus, we have optimized the hyper-parameters for the neural network:

m = 40 and λ = 0.4

Hidden Nodes vs. Accuracy

The graph of number of hidden nodes vs. accuracy for λ =0.4(optimum) is shown below. The accuracy of neural network increases with increase in the number of hidden nodes as is expected. This can be attributed to the fact that the hidden nodes represent the number of learned features. Also, we can observe from the graph that after a certain value of hidden nodes, the accuracy will become constant. This is because the accuracy cannot increase beyond a fixed value, since the number of features to learn is also fixed.



Hidden Nodes vs. Training Time

The graph shown below depicts the relation between the number of hidden nodes (m) and the total training time(on our local system). The training time increases as the value of m increases. This is because the number of weight calculations and updates increases exponentially for each additional hidden node.