CSE 574 INTRODUCTION TO MACHINE LEARNING

Programming Assignment 1

Handwritten Digits Classification Using Neural Networks

Team# 35

Bhavani Sundara Raman – 50169253

Prithvi Ganesan – 50169094

Sivasathya Pradha Balamurugan - 50169073

Introduction

Our task is to implement a Multilayer Perceptron neural network and evaluate its performance in classifying handwritten digits. We achieve this by performing tasks that are in modules like pre-processing, ObjFunction and predict functions.

Implementation:

Preprocess – First we load the test, train and validation data sets. We then split the training dataset into 50000 and 10000 as training and validation datasets respectively in a random fashion. Test, train and validation labels are generated from the respective data sets. Feature selection is done in order to eliminate from the data sets those inputs that contain no values or uniform values for pixels. This function returns the labels as vector.

ObjFunction

- **Feedforward –** The input after multiplying with weights is passed through the hidden layer and sigmoid operation is performed for that node. Similarly the result from the hidden layer is passed to get the output from the output layer after multiplying with the weights.

- **Backpropagation –** The error function is calculated and derivative of error function with respect to both the weights are calculated.

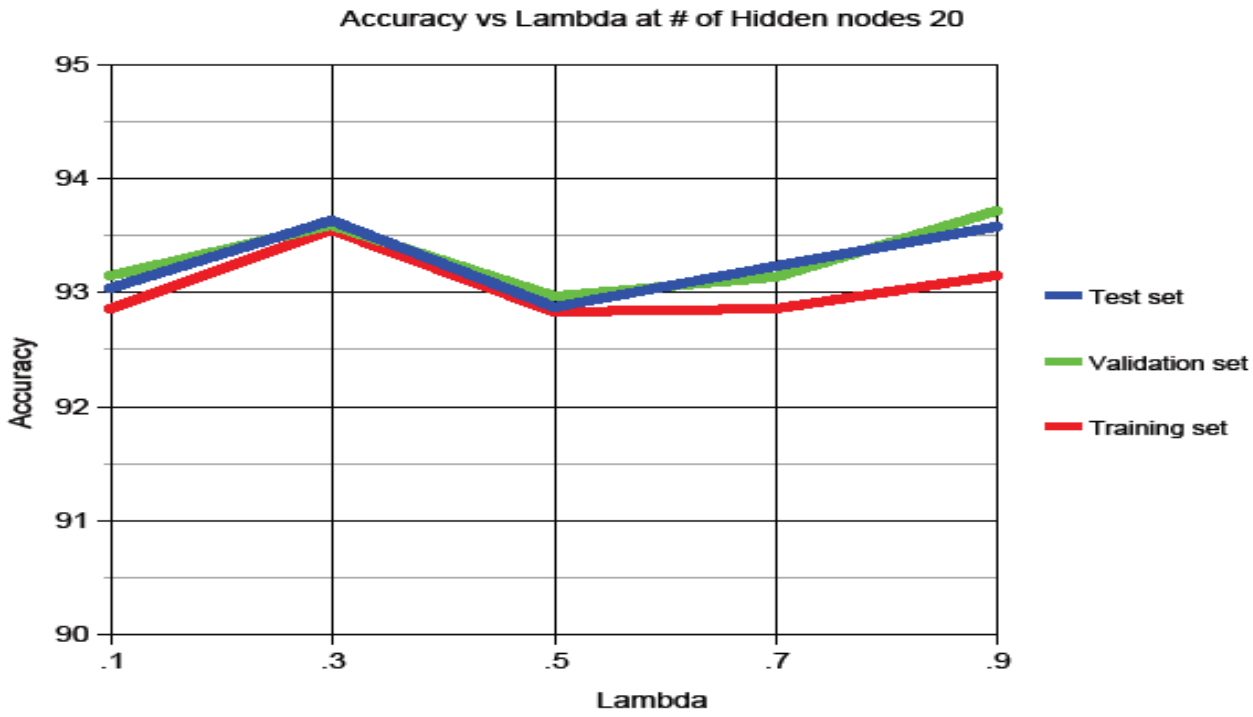
- **Regularization –** Using the given equation we compute the regularization term and add it to error function in order to limit the parameters. This helps in avoiding the overfitting and underfitting problem. The gradient of the weights are calculated.

Predict – Feedforward is done in order to predict the labels.

Choosing hyper-parameters – Observations made:

Accuracy vs. Lambda

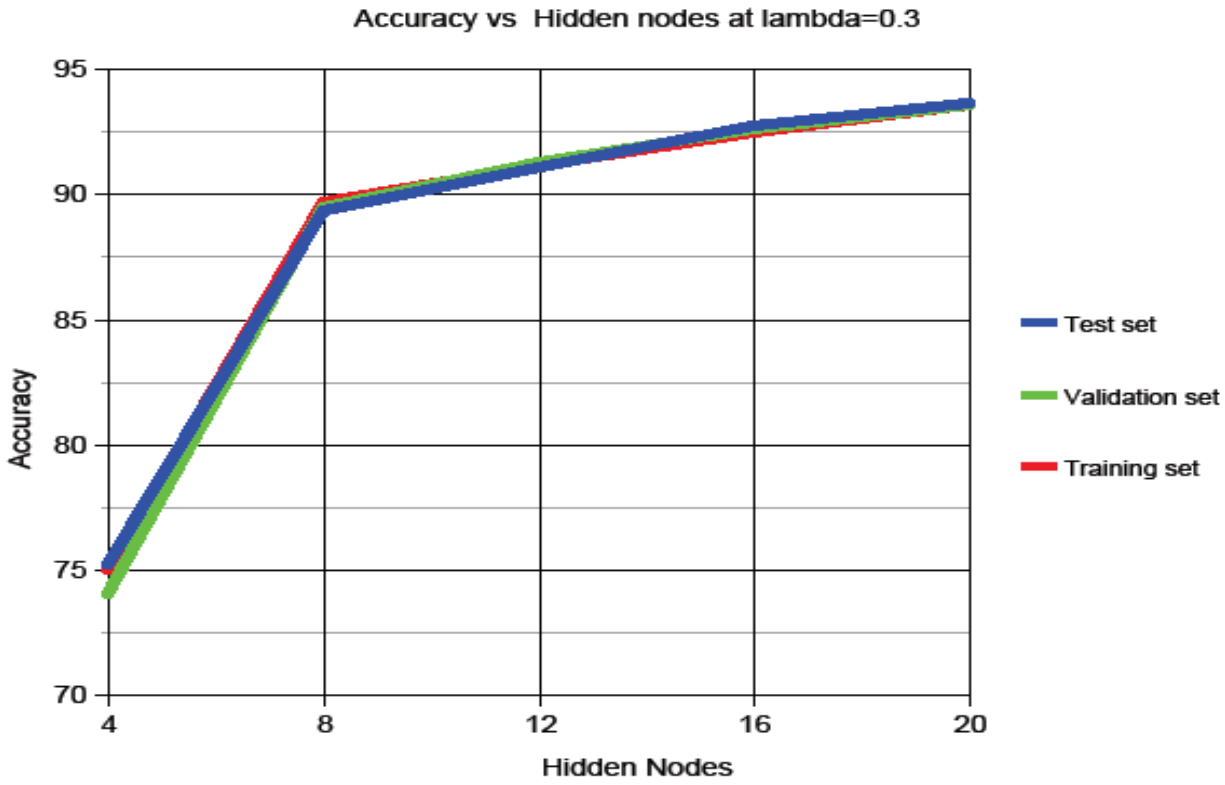
Lambda value can significantly affect the accuracy in predicting the training set and validation set results. The regularization value (i.e. lambda) should be not too high or low as it would either give poor generalization (overfitting) or doesn't find patterns in the training set and test set (underfitting).



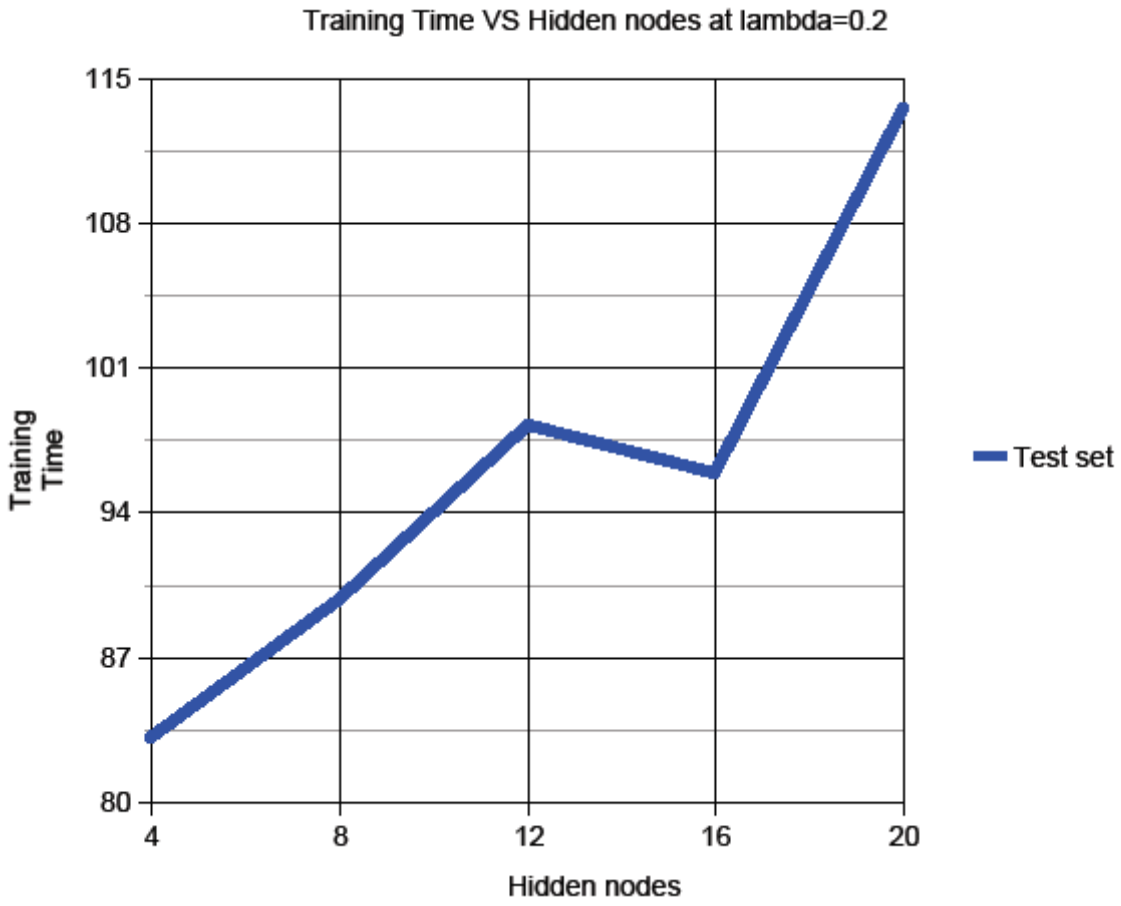
The optimal (hyper parameter) number of hidden nodes is selected as 20, as it gives us the best accuracy amongst other values of 4,8,12, 16 and 20.

Accuracy vs. Hidden units

We vary the number of hidden layers and lambda values to study the accuracy in each case. From the various cases we have identified the accuracy to be the optimal when lambda value is **0.3**. We achieve the maximum accuracy of **93.54%** with this than in other values of lambda between 0 and 1 in increments of 0.1.



We can infer from the above that with increase in number of hidden nodes, the accuracy of the data set keeps increasing. But increase in the number of hidden nodes means more number of features to learn for the neural network. Thus, when the complexity of the neural network increases, the training time also increases (as shown below).



Conclusion:

By varying the values of Lambda and the number of hidden nodes, we find the optimum value of the hyper parameters. We have achieved maximum accuracy of **93.54%** with **20** hidden nodes and **0.3** as lambda value in classifying the test data.