CSE 574 Programming Assignment 1
Handwritten Digits Classification

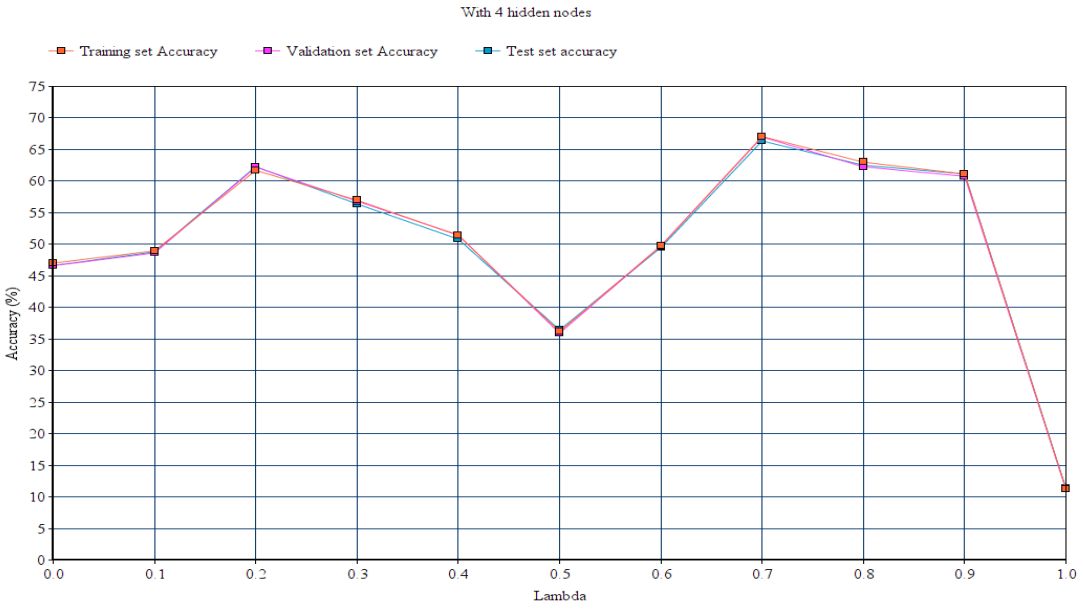
Group number 12
Bhavika Jain - 50170168
Chen Song - 50060333
Pranav Jain - 50169659

1. Approach

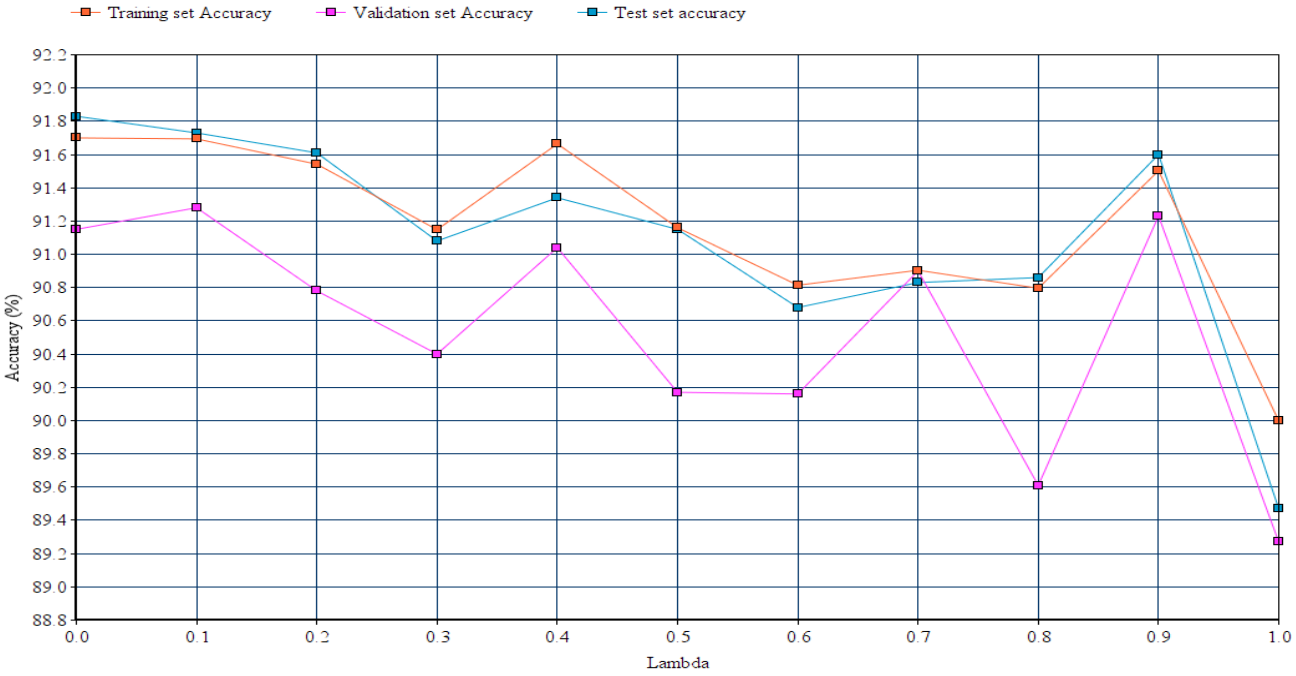
We conducted experiments on our neural network for various values of hyper-parameters and hidden units in order to determine the influence of these parameters on the network performance. We started with 4 hidden units and varied lambda from 0 to 1, in increments of 0.1. We increased the number of hidden units to 8, 12, 16, 20, 50 while continuing to vary lambda in same pattern. We measured training, validation and test set accuracy and network training time for all these cases and plotted them against different hyper-parameters and hidden units. Based on the results, we selected the parameters that give best performance. In the below section, we have plotted our observations and some of the most significant plots derived from our data.

2. λ vs. Neural Network performance (Accuracy %):

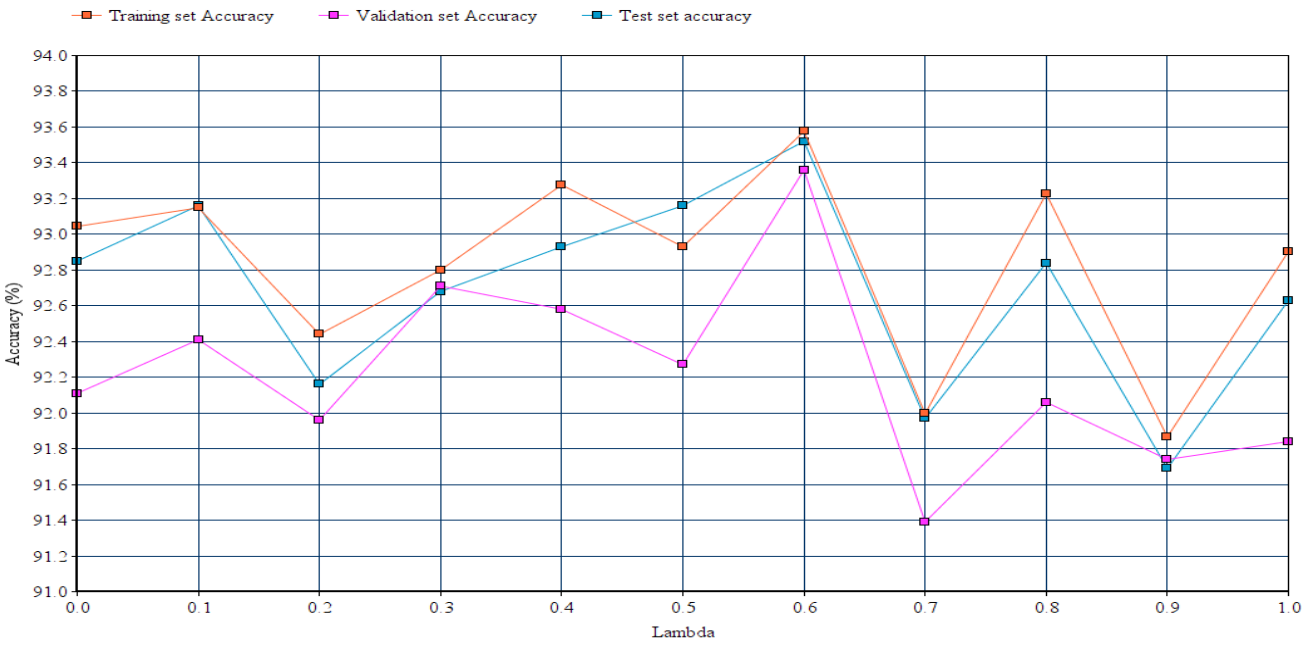
The graph below shows the performance of our neural network for different values of lambda - from 0 to 1 for different number of hidden units. With increase in lambda values, we expected the training set accuracy to decrease and test set accuracy to increase. However this was not true for all lambda values as shown in the plots below. The reason could be due to similarity of test set and training set. It could also be due to random selection of training samples and initial weights. Overall, we noticed that there is no significant improvement in performance due to lambda value changes when the number of hidden units is high.



With 12 hidden nodes



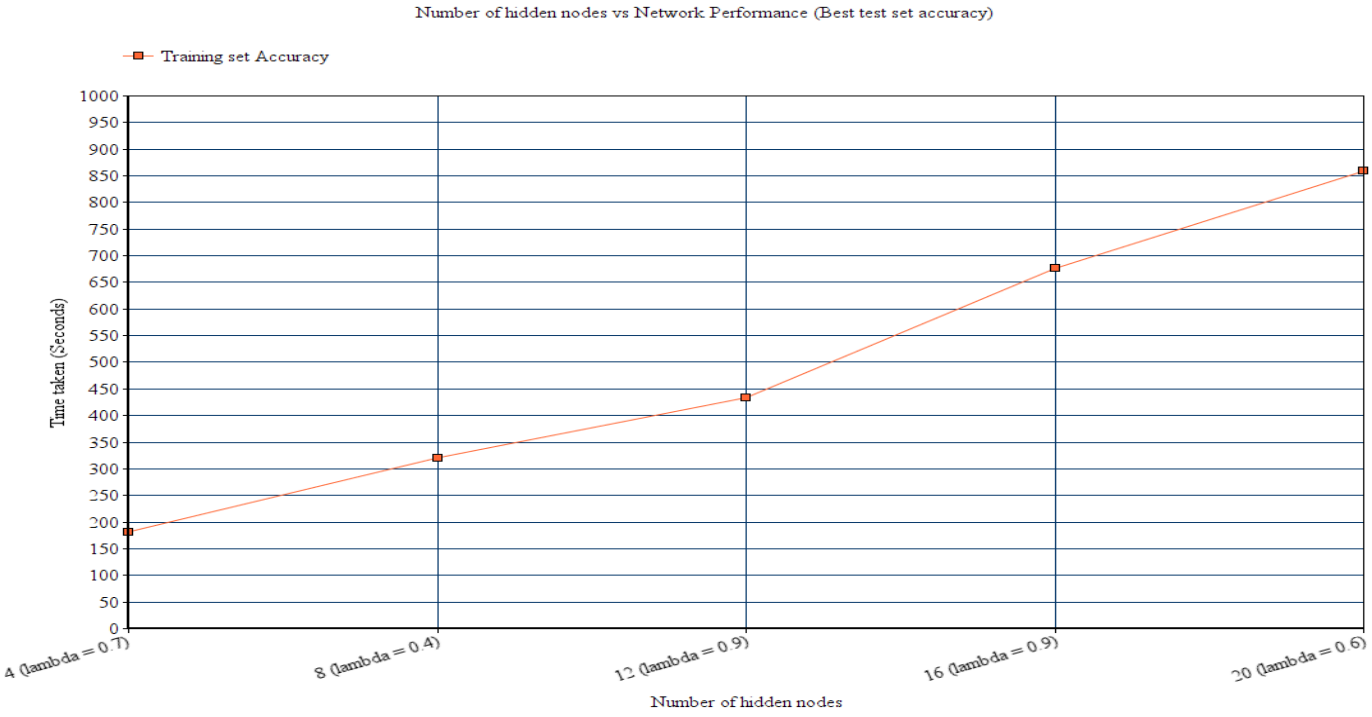
With 20 hidden nodes



3. Relation between number of hidden units and training time:

We observe that as the number of hidden of units increases, the training time also increases. The reason being more hidden units implies more weight and gradients to compute. Consequently, the computational complexity of the network increases and also more time is taken to converge to an optimal solution.

In the below graph, we have plotted the time taken by the neural network against different number of hidden units and lambda values for which we got maximum accuracy for the training set.



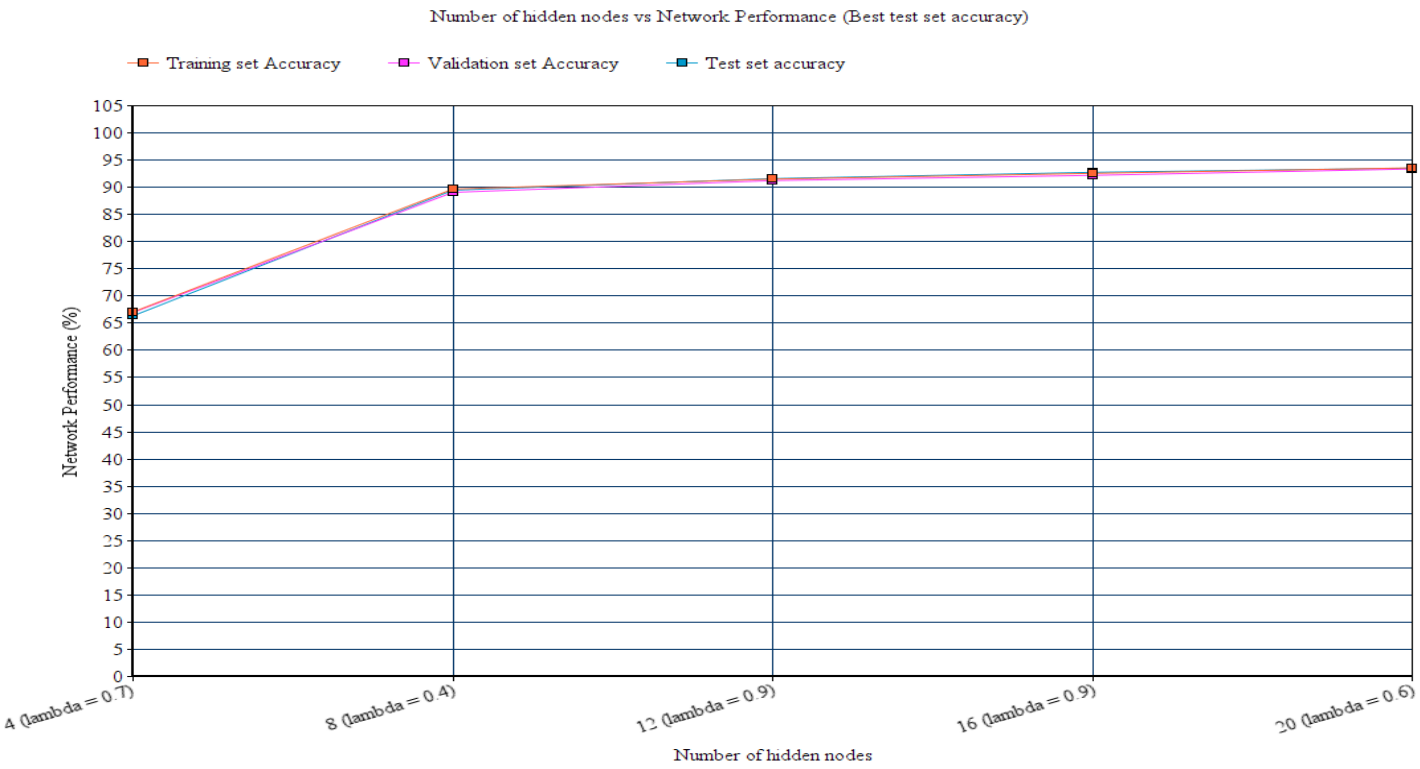
4. Number of hidden nodes *vs.* Network performance (Accuracy %) :

The graph below shows the values of hidden units and lambda for which our neural network provided the maximum accuracy % for each data set (Training, Validation and Test)

As we increase the number of hidden units, the performance of our neural network improves. This is an expected behavior as large number of hidden units provides higher precision.

As shown in the graph below, the performance of the neural network is poor when the number of hidden units is less. However as the number increases, the performance improves and after a while gets steady.

With 50 hidden nodes and lambda 0.6, the training, validation and test set accuracies were 93.764%, 93.74% and 93.88%. It is clear that after a certain point, the accuracy almost remains constant. In this model, it stays around 93% for all the sets.



4. Conclusion:

As shown by the plots, we tried different number of hidden units and lambda values in order to determine the most optimal setting. As we increase the number of hidden units, the accuracy percentage increases. If time is not an issue, the best number of hidden units is greater than or equal to 20 in this scenario, as shown in the graph above. Beyond this, the accuracy saturates as we increase the number of hidden units. We tested our network performance up to 50 hidden units and there was no major improvement in the accuracy. As stated in the section above, the accuracy hovers around 93% even if the hidden node number is increased to 50 with lambda value 0.6. So, if time is an issue, 8 to 12 hidden nodes seem to be a good choice.

The optimal value for lambda changes as the number of hidden units changes. By inspecting the plot above, for a network with 20 hidden units the optimal value is 0.6 for which the training, validation and test accuracy are all in the range 93-94%.