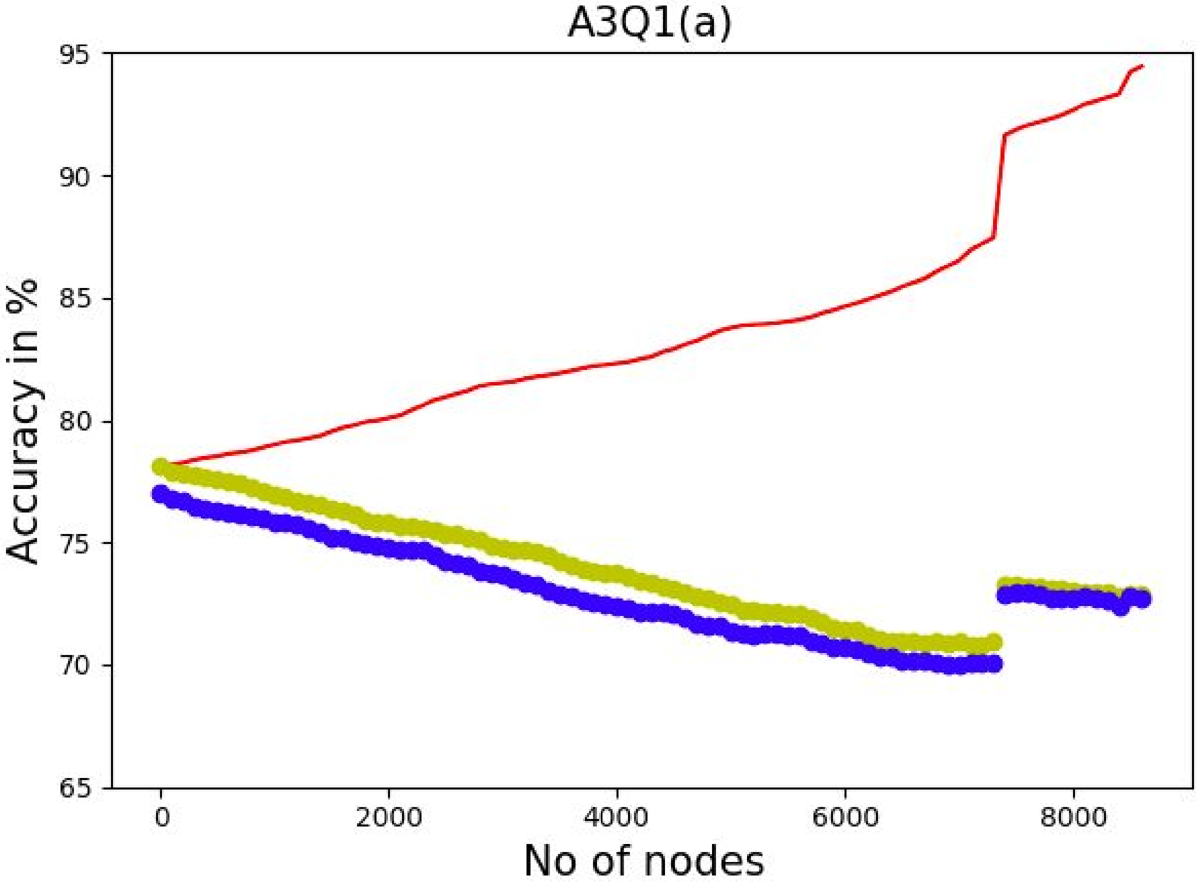
Assignment 3 Report
COL774
Shivanshu Verma
(2015CH70186)

Question 1 : Decision Tree

a)



Red: Training set accuracy
Yellow : Test Set Accuracy
Blue : Validation Set Accuracy
Choice of Node : DFS growth

Here as the majority prediction was large, we can see a decrease in accuracy. Also, in DFS, we go in the depth first which means it will only predict better when the particular column value is what in the node split value and for the rest , it will predict the value at the same node no matter how deep the tree has grown.

The discontinuity in the curve is when all the children of first node are created and now only depth of last node is remaining. This is because, we have option for every value of particular column.

b)

c)

d) Parameters:

Default values:	71.7 %
Max_depth = 2 :	80.35 %
Max_depth = 5 :	80.18 %
Min_sample_split = 1 :	71.7 %
Min_sample_split = 50 :	76.8 %
Min_sample_split = 300 :	79.7 %
Min_sample_split = 1000 :	80.3%
Min_sample_split = more than sample :	77 %(Majority Prediction)
Min_Sample_leaf = 10 :	80.23 %
Min_Sample_leaf = 50 :	80.1 %
Max depth =5, min_sample_split = 50, min_sample_leaf =10, criterion = ‘gini’	80.26%
Max depth =5, min_sample_split = 50, min_sample_leaf =10, criterion = ‘entropy’	80.38%

For this dataset, max_depth = 5, min_sample_split = 50 and min_sample_leaf = 10 were found to be optimum with entropy criterion whereas using other things default, but criterion ‘gini’, it was observed that gini is better criterion than entropy in terms of accuracy. The reson behind this optimal parameter is all about the overfitting of decision trees, we need to stop growing tree at a limit to avoid overfitting.

Optimal Parameter : criterion = ‘entropy’, max_depth = 5, min_sample_split = 50 and min_sample_leaf = 10

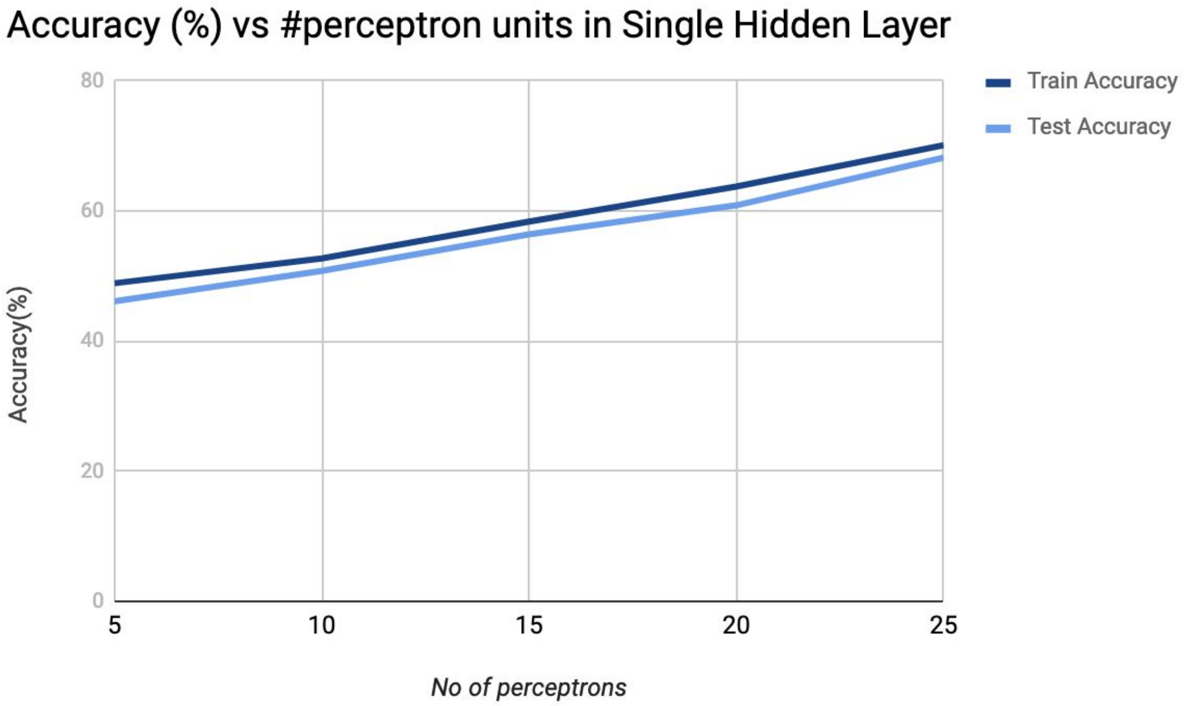
e) After one hot encoding, accuracy on validation set for optimal parameters of part d is 79.86%
As the no of variables increased by one hot encoding, depth must be reduced to get optimal parameters for high accuracy.
Optimal Parameters for this are : criterion='entropy', max_depth=3, min_samples_split=50, min_samples_leaf=10, givaing a accuracy of 80.17%

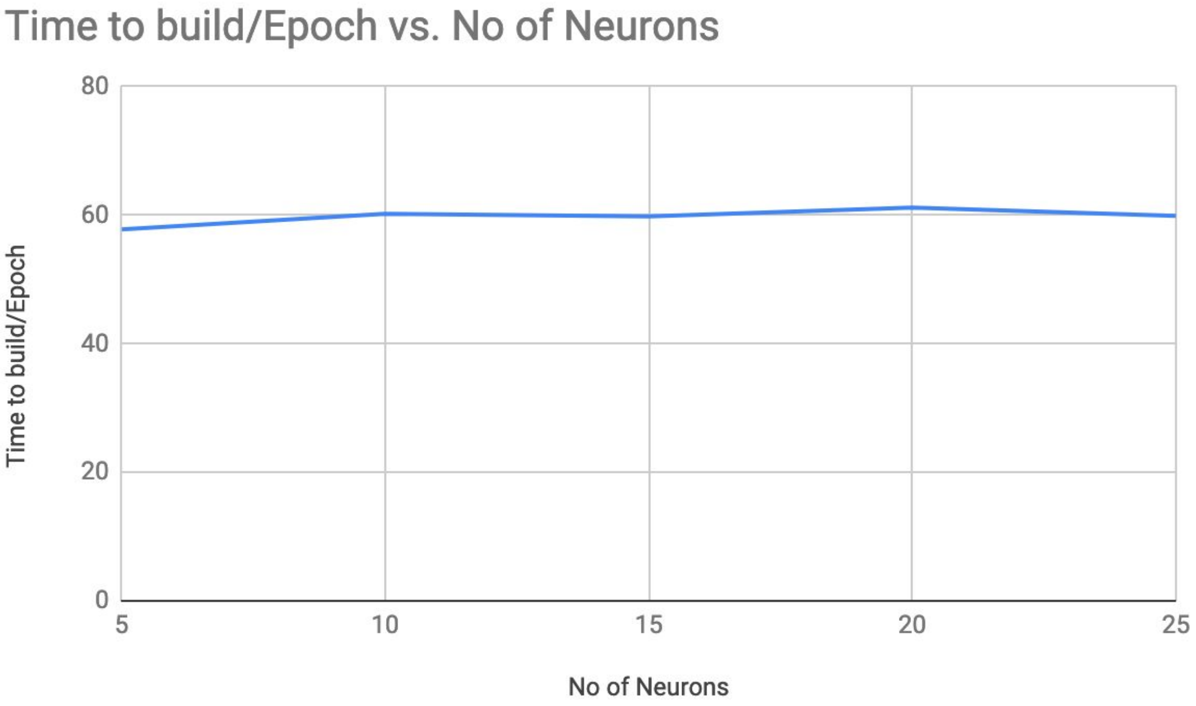
f) For default values : Accuracy = 78.9%
For n estimator to be 100: Accuracy = 80.2 %
For n estimator to be 1000: Accuracy = 80.15 %
i.e increasing no of trees after a limit won’t help.
As same one hot encoding here also, using previous optimal solution, with optimal n_estimator, Accuracy = 79.05%

Trying different combinations, optimal was found when max_depth was none, min_sample_split =10 and min_sample_leaf= 5 which is 80.33%. This is because, due to avergae of all the random trees, the problem of overfitting does occur late than using only a single tree. So we can have little relaxation on depth and other parameters.

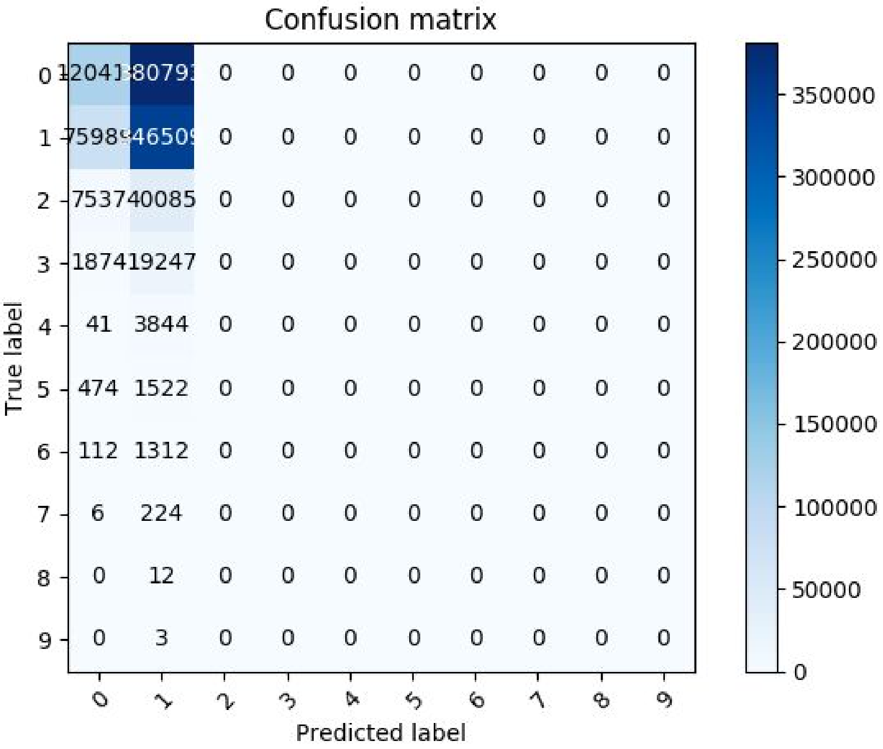
Question 2 : Neural Networks

a) Done
b) Done
c) Graph for accuracy is:

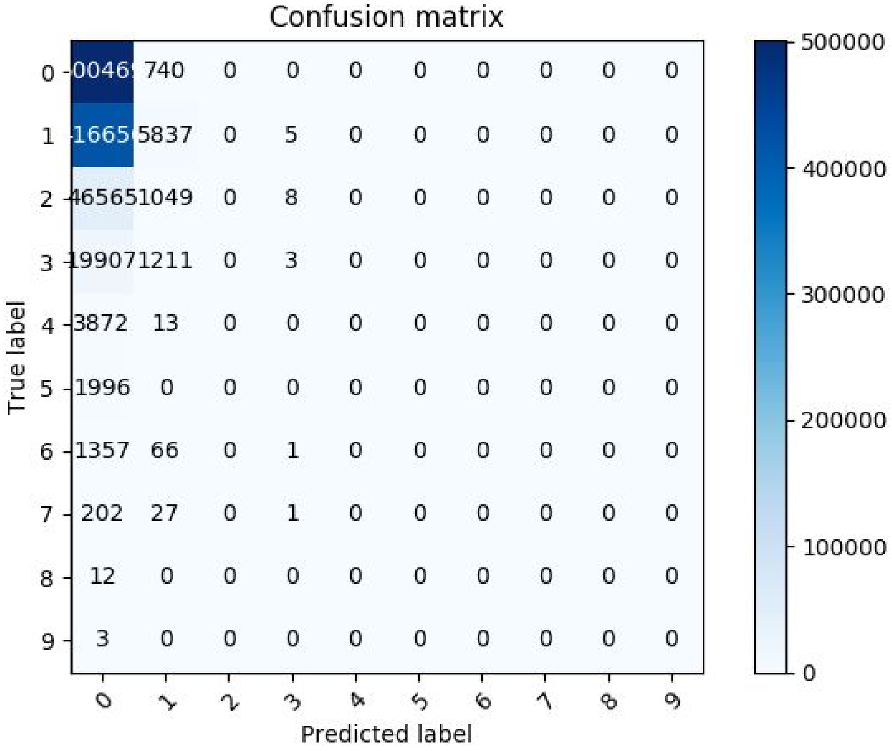




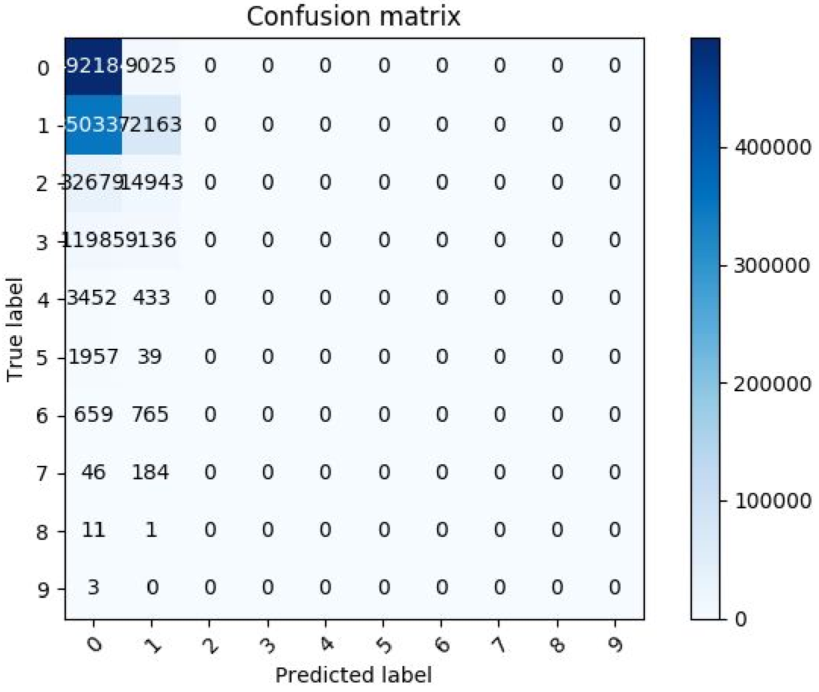
Confusion Matrices are :



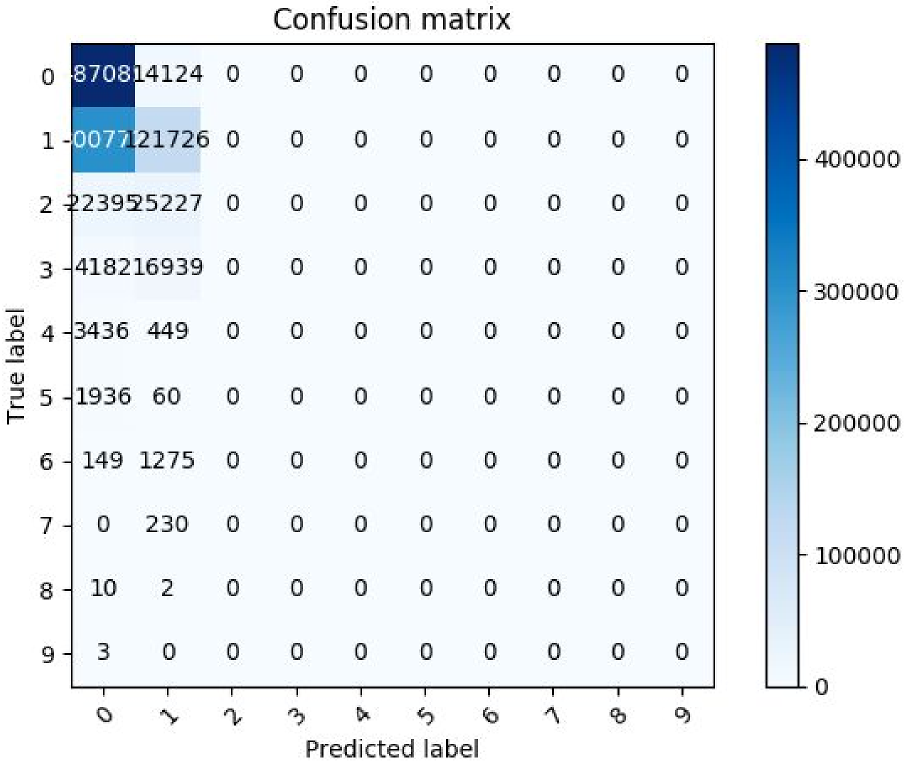
For 5 unit perceptrons with 46.12% test accuracy



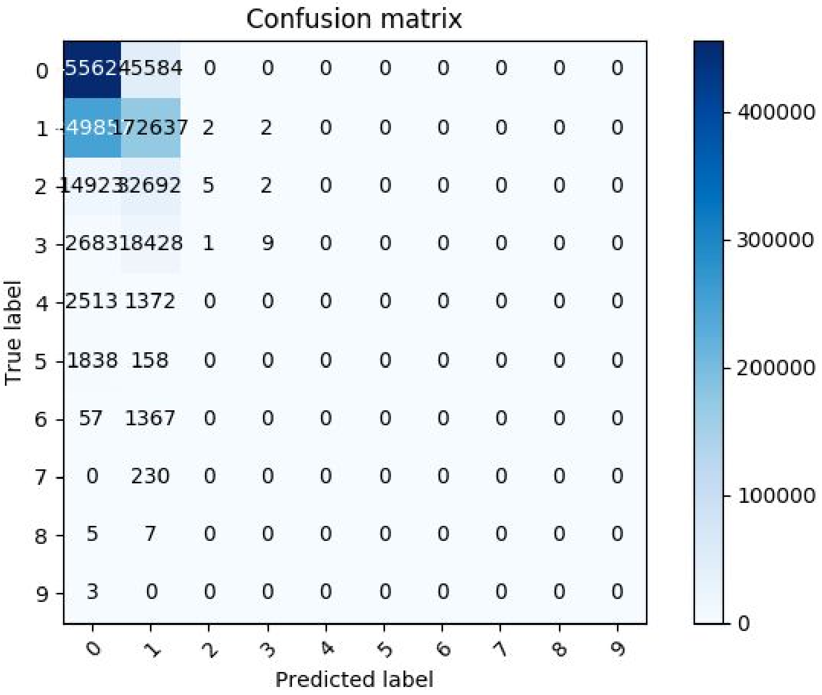
For 10 unit perceptrons with 50.81% test accuracy



For 15 unit perceptrons with 56.43% test accuracy



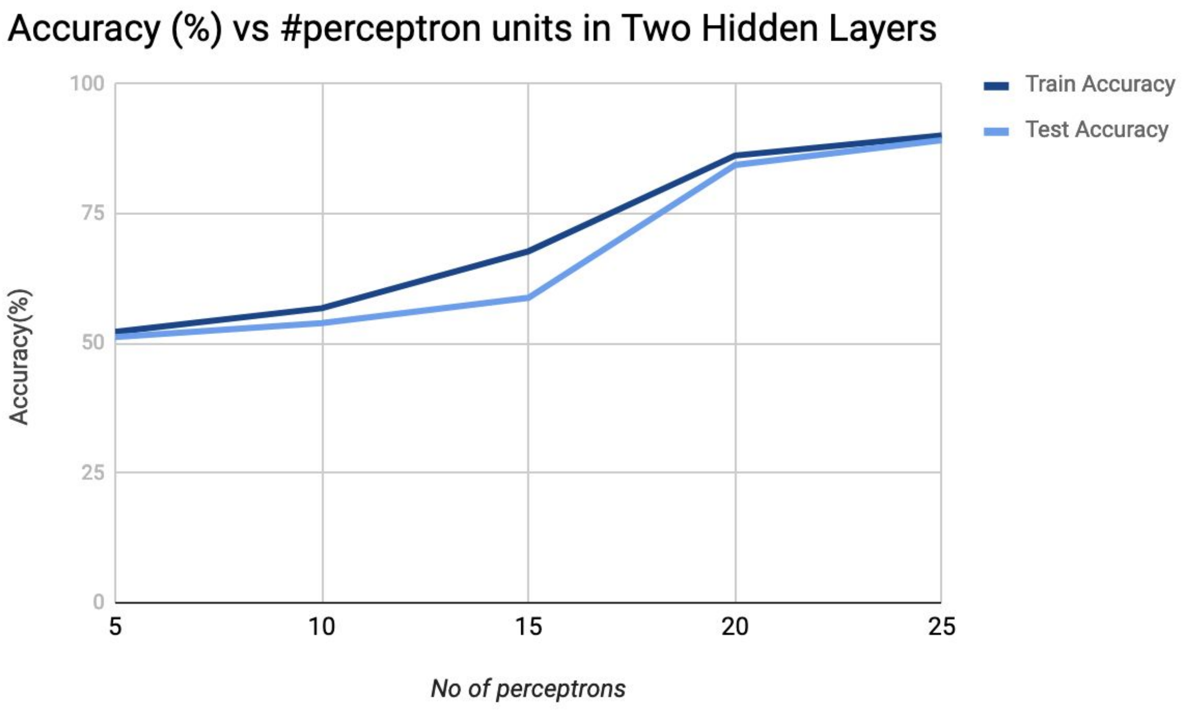
For 20 unit perceptrons with 60.88% test accuracy

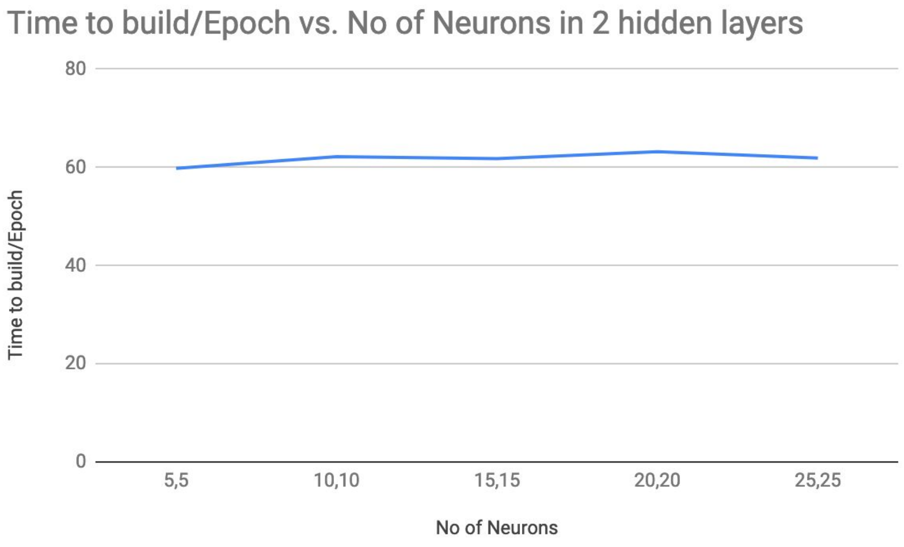


For 25 unit perceptrons with 68.22% test accuracy

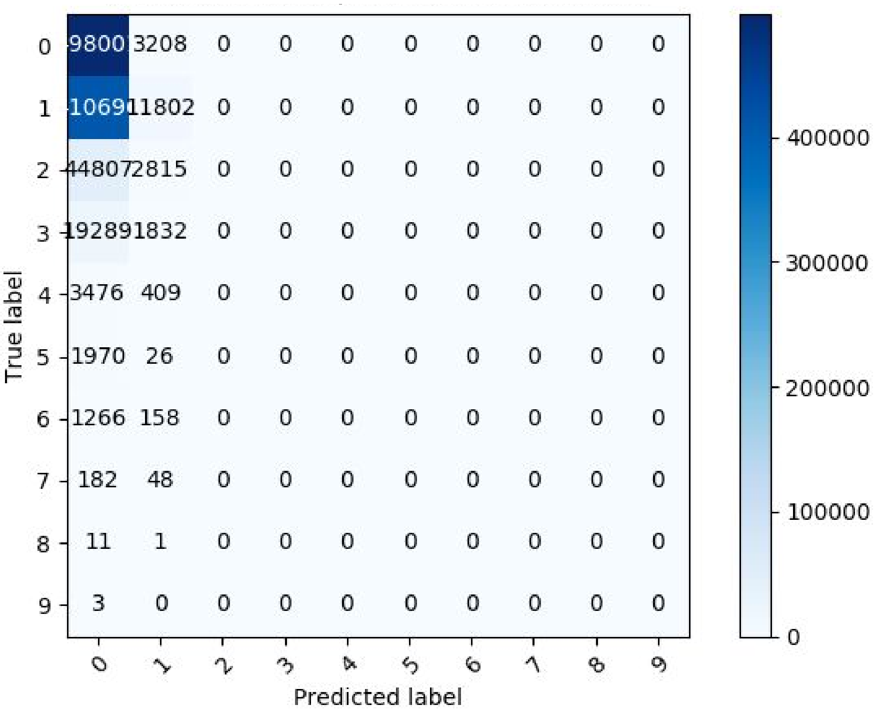
In this part c), it is observed that the accuracy for both the training set and test set increases with the increase in no of perceptron units in a hidden layer. Therefore, in confusion matrix, we can observe the movement of dark shaded part from side to the diagonal part as we increase the no of perceptron units within a hidden layer. Max achievable accuracy was 70.12% with 25 perceptron units.
Also for time to train, for me it is taking approximately an hour to train when training with 100 epochs, whereas the time per unit epoch is plotted above and is somewhat similar for all.

d) Graph for accuracy is:

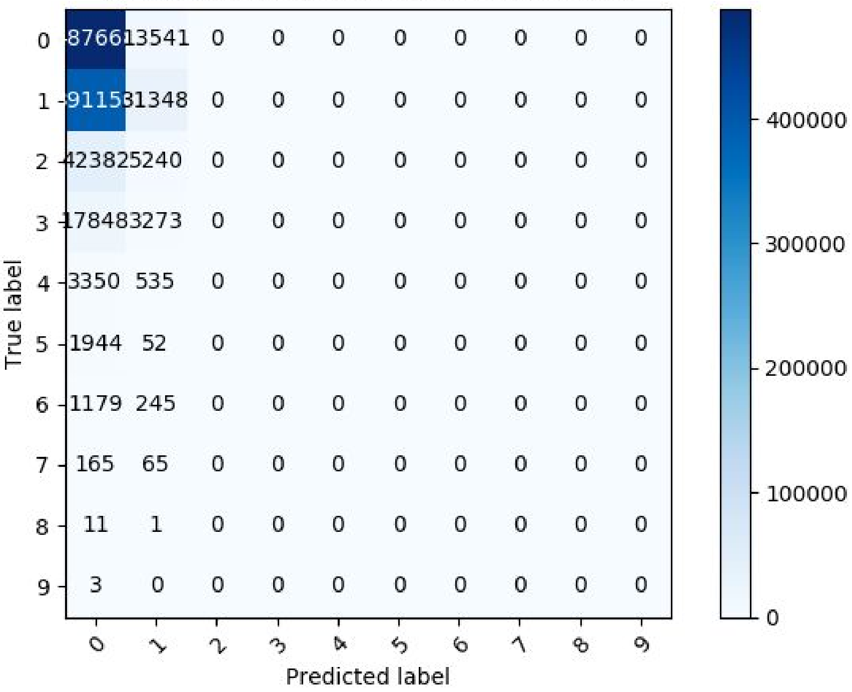




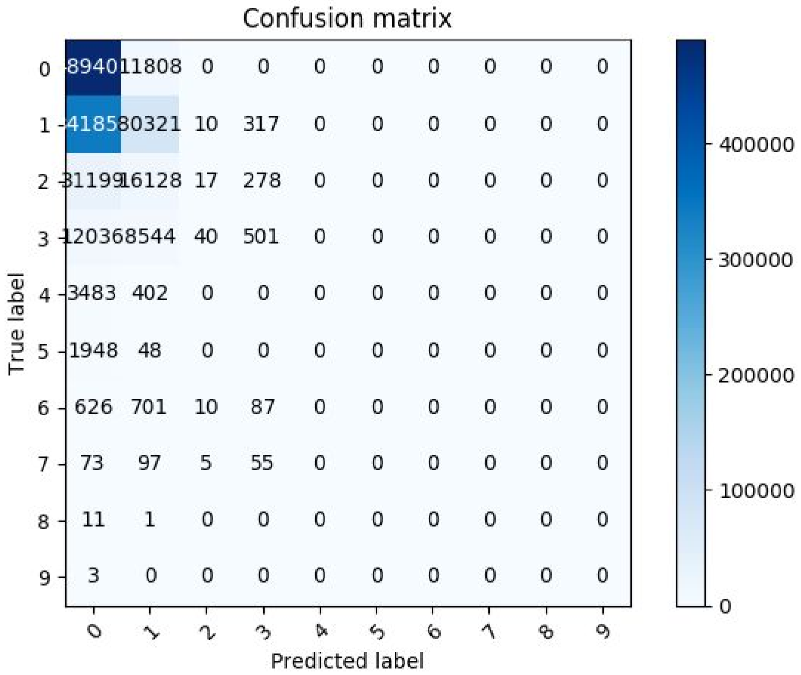
Confusion Matrices are



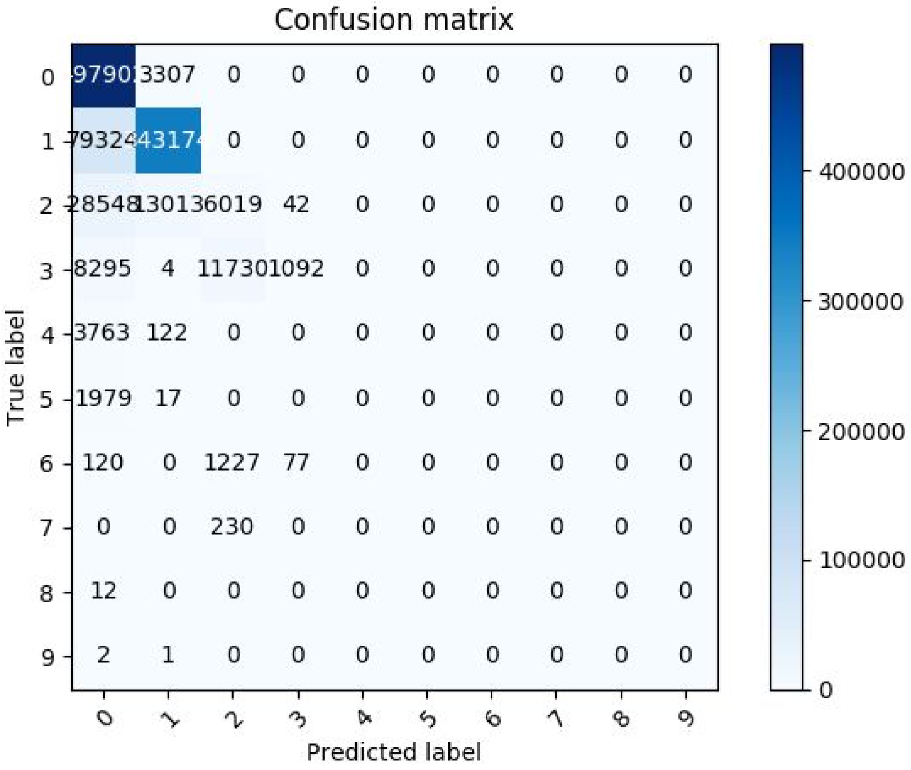
For 5,5 unit perceptrons with 51.12% test accuracy



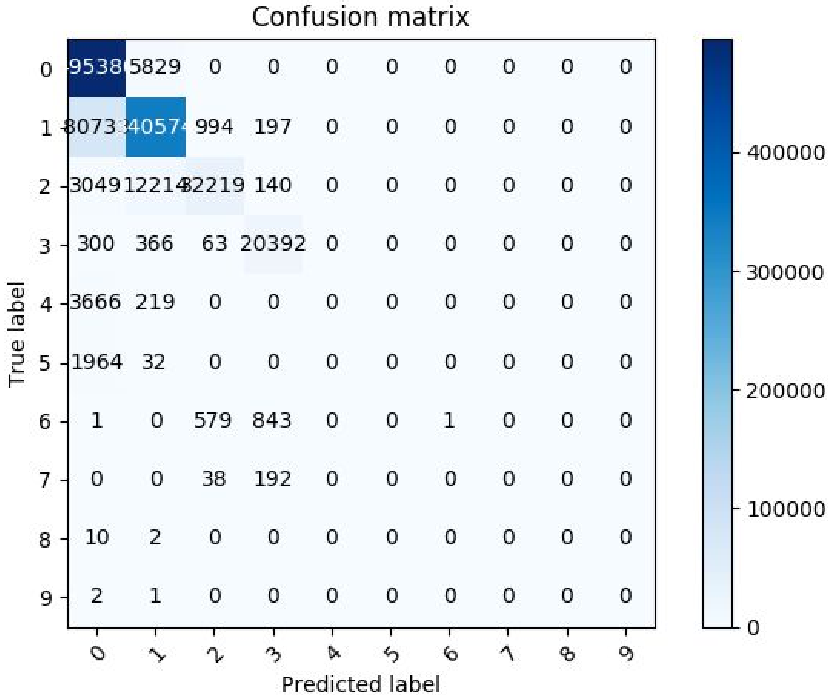
For 10,10 unit perceptrons with 53.91% test accuracy



For 15,15 unit perceptrons with 59.5% test accuracy



For 20.20 unit perceptrons with 84.8% test accuracy

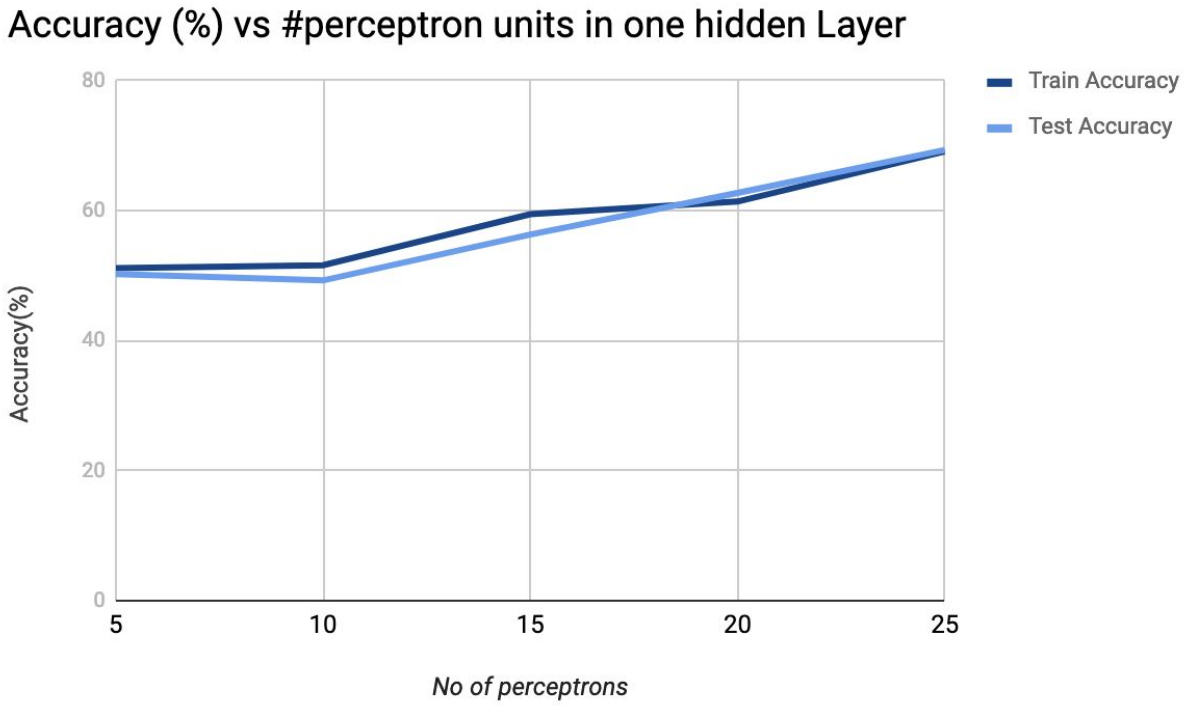


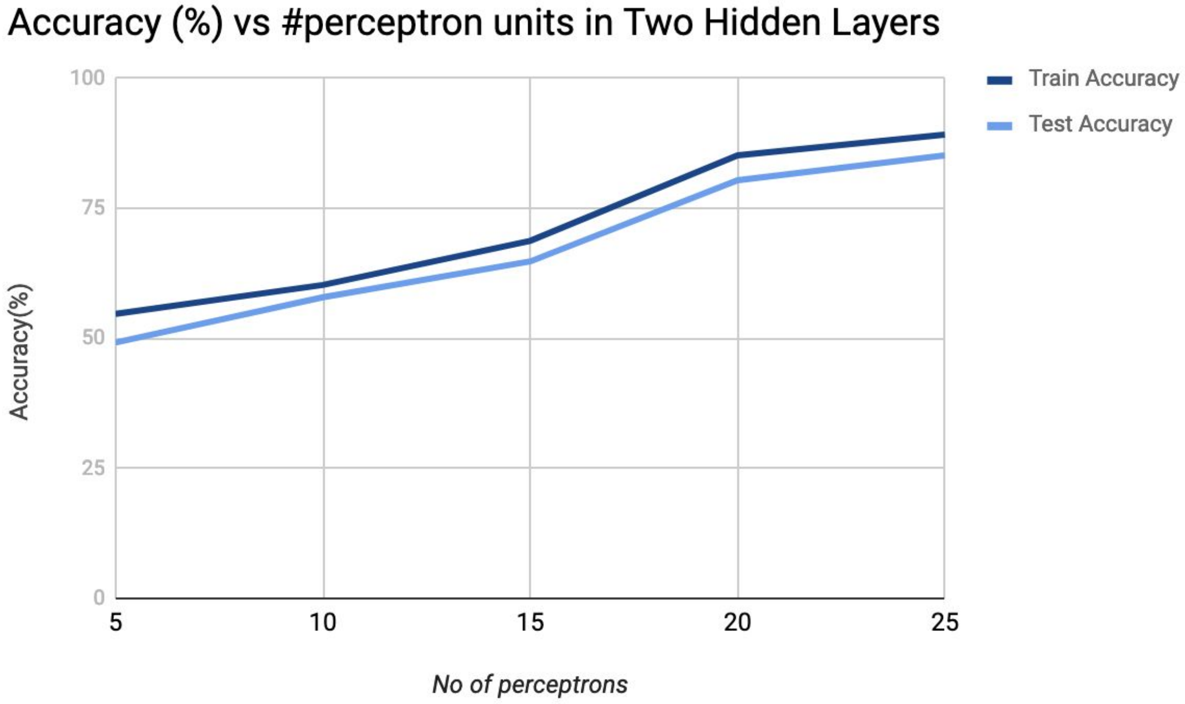
For 25,25 unit perceptrons with 89.12% test accuracy

In this part d), it is observed that the accuracy for both the training set and test set increases with the increase in no of perceptron units in two hidden layers. Therefore, in confusion matrix, we can observe the movement of dark shaded part from side to the diagonal part as we increase the no of perceptron units within a hidden layer. Here from part c , the difference in accuracy can be seen easily. The maximum achievable accuracy for this part was 89.19% whereas for the part c was 70.12 %
Also for time to train, for me it is taking approximately an hour to train when training with 100 epochs, whereas the time per unit epoch is plotted above and is somewhat similar for all but higher than part c where only one hidden layer was there.

e) In this part, accuracy found for all the perceptron units was little higher as that of in part c) and part d). This was because, when we use adaptive learning rate, we decrease the parameter whenever there is a oscillation between two values and training error does not decrease with more than some tolerance. Otherwise rest results are same for this part too, confusion matrix can be observed to move darker region towards the diagonal as the accuracy increased with no of perceptron units.

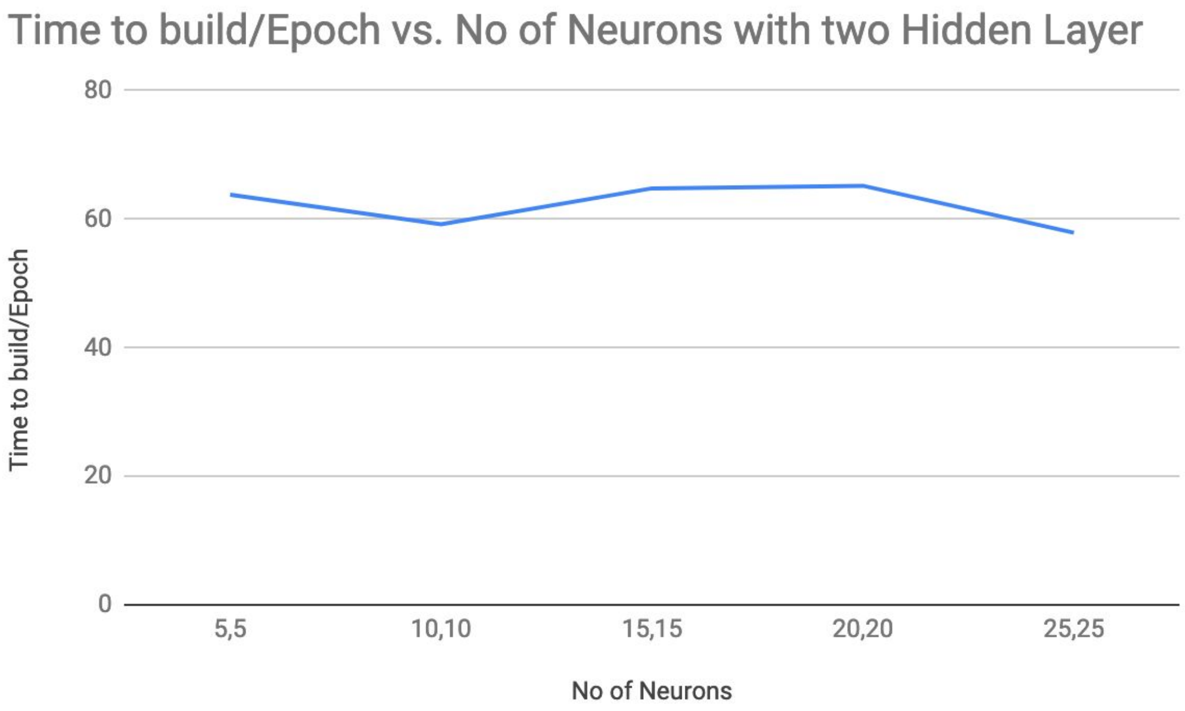
Graph for accuracy of single hidden layer:



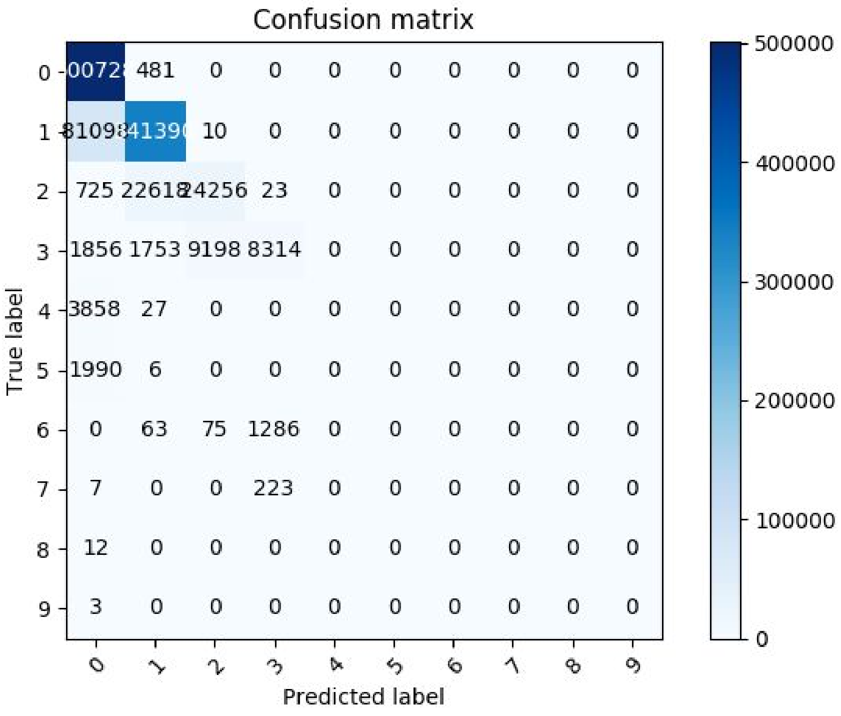


In this case two, the accuracy was just slightly more and little uncertain as somewhere increased and somewhere decreased than that of part c and d. Max accuracy achievable was again for 25 and 25,25 which were 69.34 and 87.42 respectively.

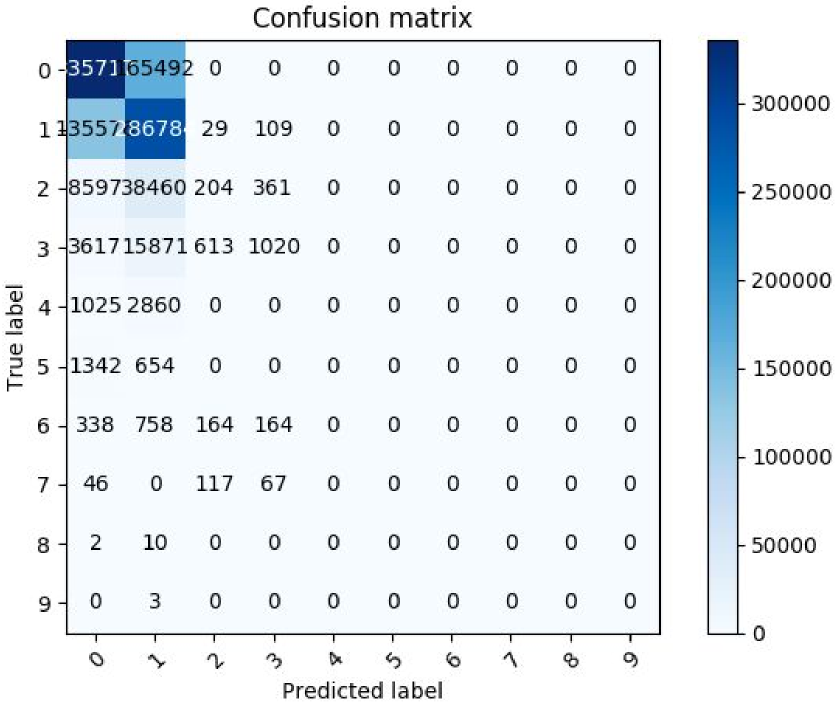
For the time plot, ideally it should be little lower and accuracy should increase as learning rate is decreasing , but as I ran too many codes at the same time for results, the time graph was little higher and uncertain than those of in part c and d. Given below is the plot:



Given below are the confusion matrix for highest accuracies achieved:
For 25,25 :



For 5 perceptron single hidden layer:



f) Not Done