**Assignment 5**

**Task 1**

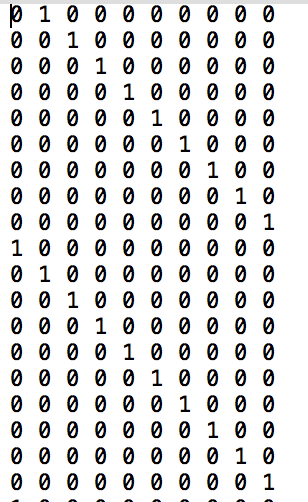
**File submitted:** Pattern.py

I have submitted Pattern.py python program which takes optdigits-orig.windep as input and gives two files as output. The first file is “first.txt” and second file is “second.txt” encoded from the given file. first.txt contains each pattern organized into single line and second.txt contains each number corresponding to unique digits of first.txt. first is the input file and second will be the target file.

Screen shot of Part of the input file:



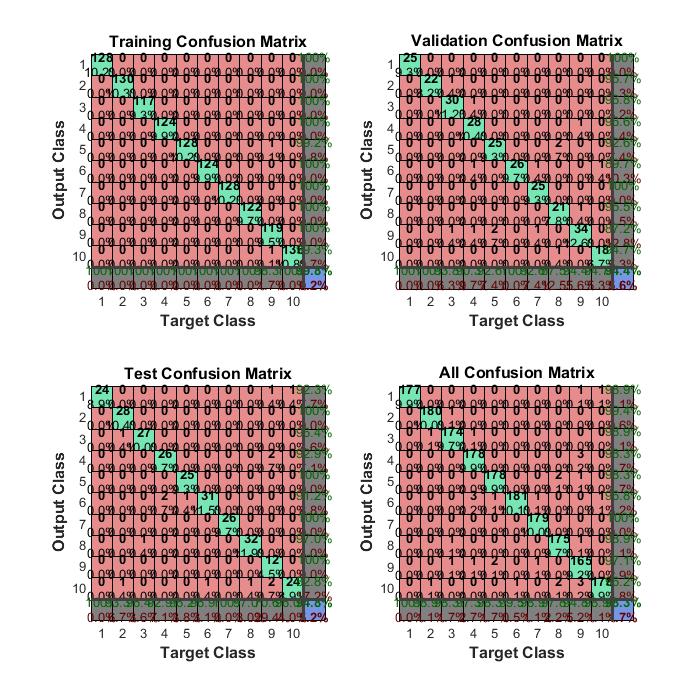
Screen shot of part of the target file:



Task 2:

I have trained the input data taken from first.txt and second.txt which are the input file and target file using MATLAB. The input and target files are feeded to the MATLAB neural network system.

Confusion plot for 10 nodes:

****

The above confusion plot is for 10 nodes when the input data(input and taget files) is given 10 neural nodes for the nntraining tool. After we train the data we get error values from which we can calculate accuracy.

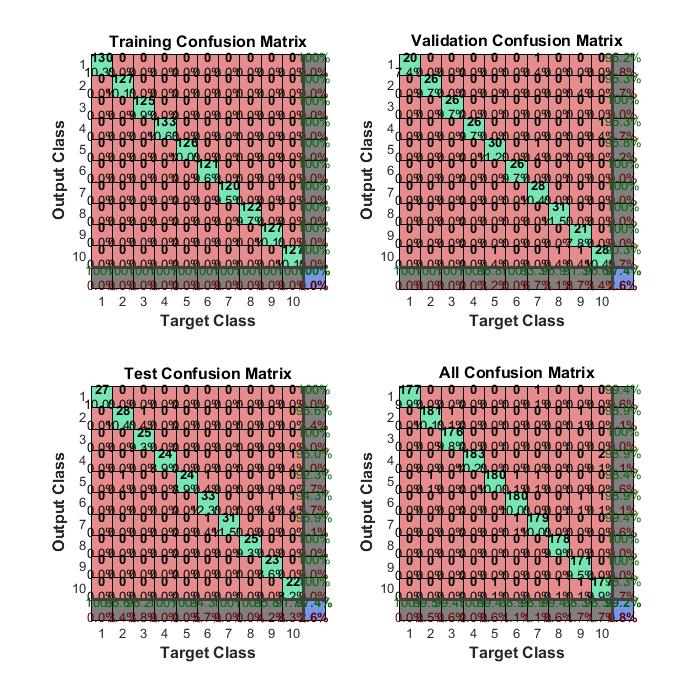
**10 nodes: Errors and Accuracy rate**

1.58982e-1 98.41018

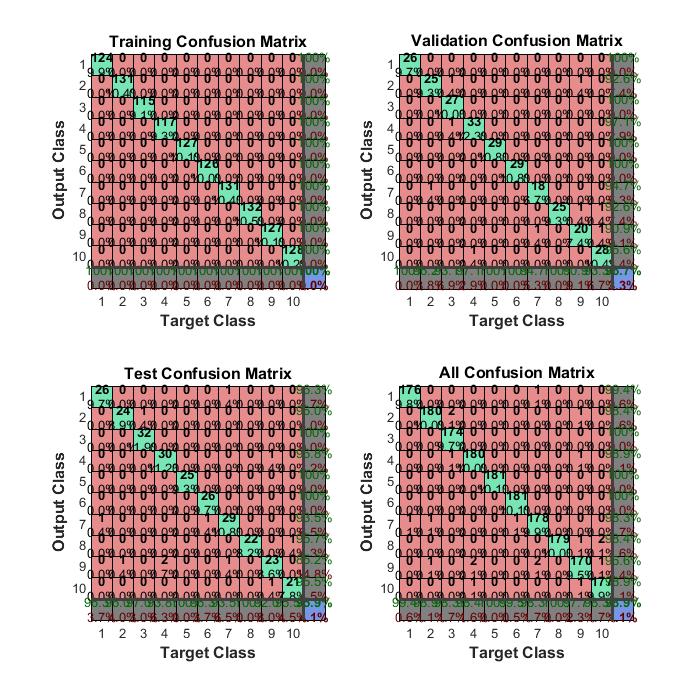
5.57620e-0 94.4238

5.20446e-0 94.79554

Confusion plot for 100 nodes:

****

confusion plot for 500 nodes:

****

**100 nodes: Errors and Accuracy rate**

0 100

2.60223e-0 97.39777

2.60223e-0 97.39777

**500 nodes: Errors and Accuracy rate**

0 100

3.34572e-0 96.65428

4.08921e-0 95.91079

**Analysis:**

Accuracy rate for 10 nodes is lesser than the accuracy rate for 100 nodes. So, increasing the nodes yield to better accuracy rate of the system. But for 500 nodes, the accuracy rate is decrease when compared to 100 nodes. Hence over fitting the nodes should be avoided.