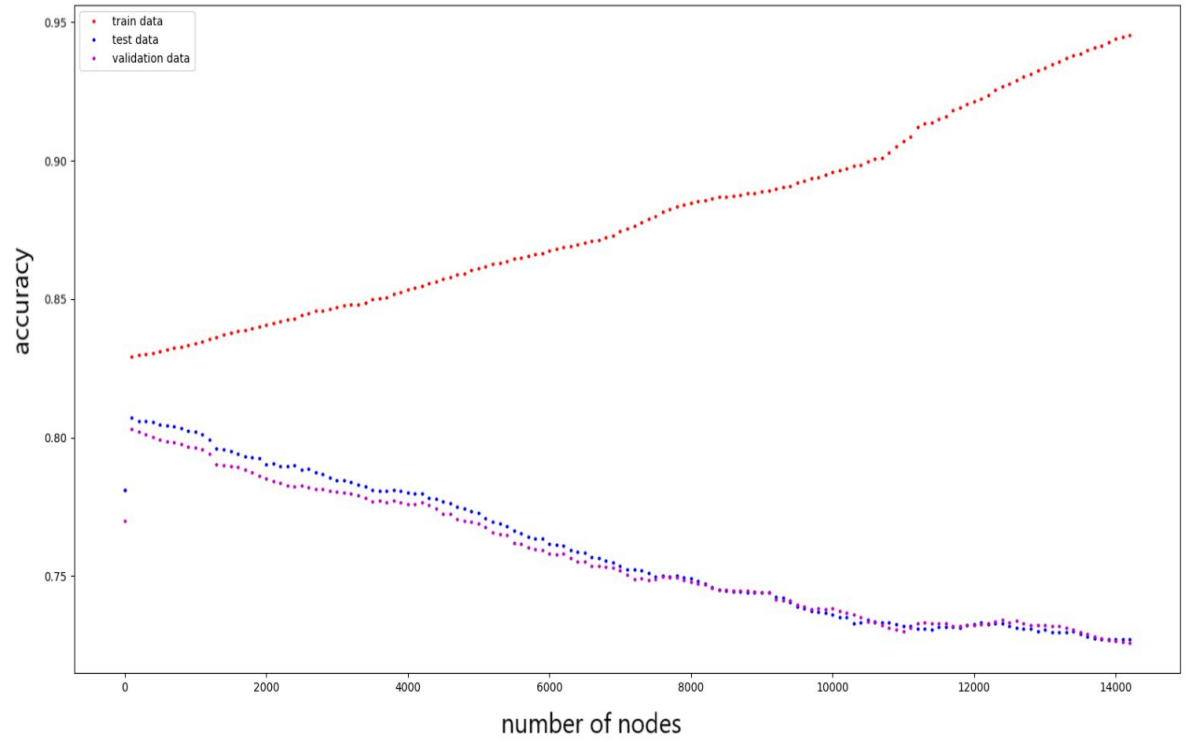
Machine Learning assignment 3

Gulshan Jangid (2014CS50285)

(PL: python 2.7)

Question 1:

(1.A)



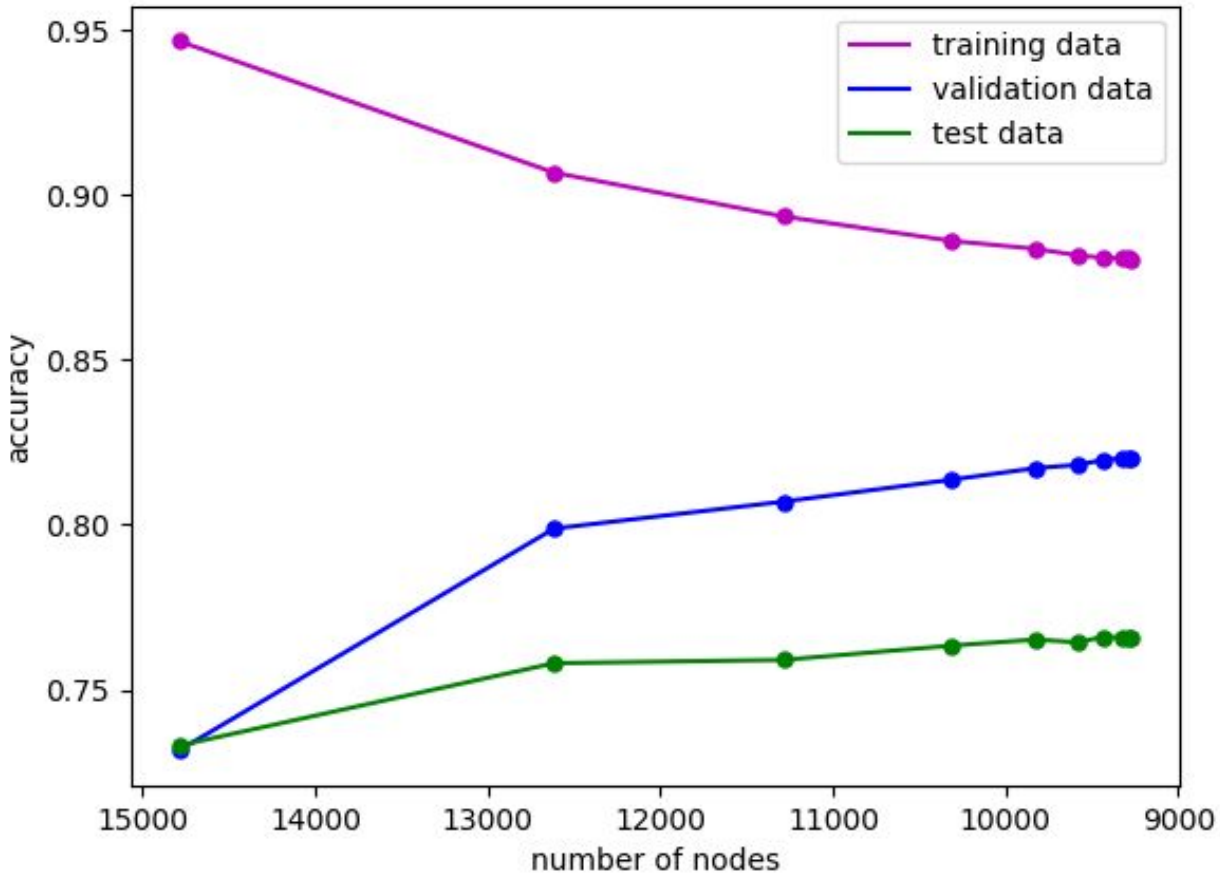
The above graph corresponds to tree being grown from left to right (Depth first search) using recursion. Growing tree takes approx **450 seconds.**

Training accuracy = 94.67 %

Validation accuracy = 73.15%

Test accuracy = 73.3%

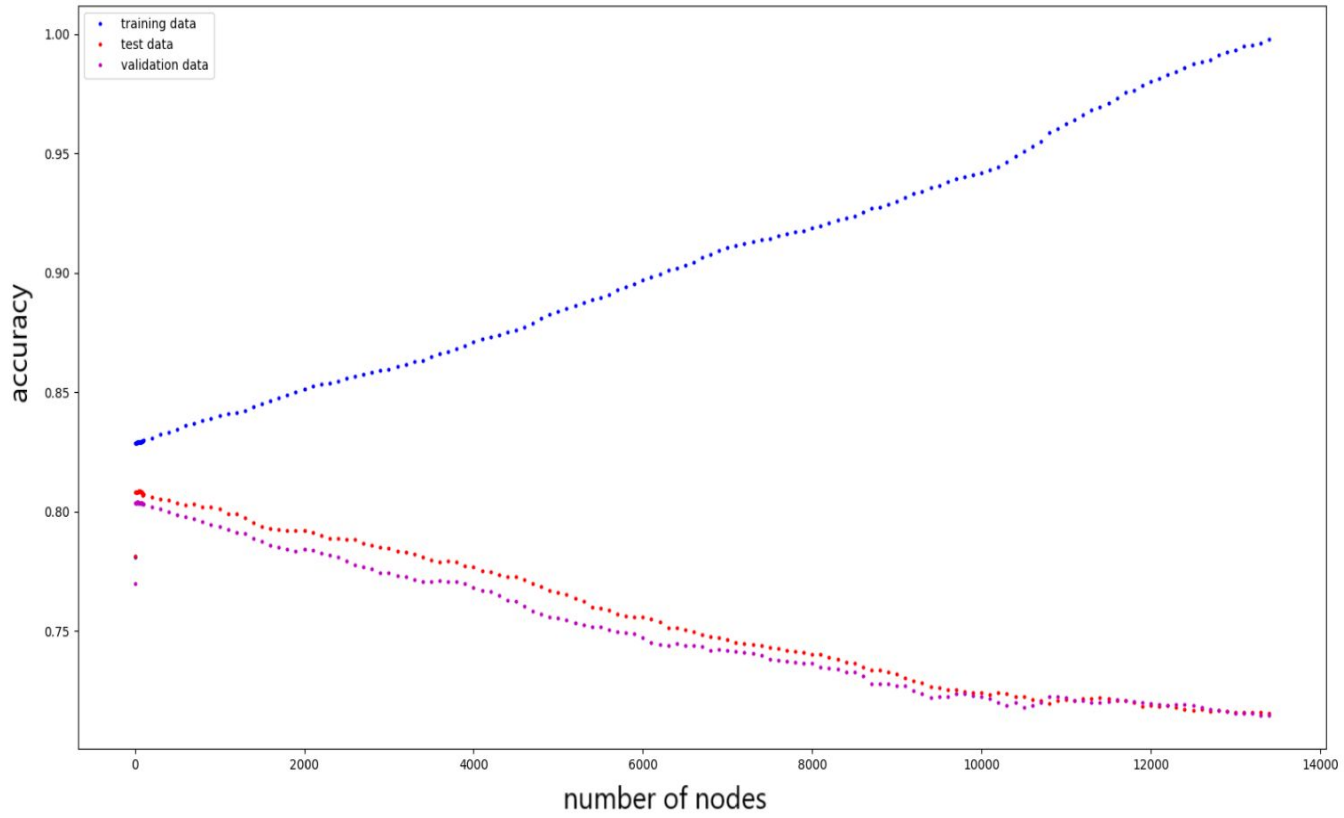
(1.B)



After pruning:
Training accuracy = 88.06 %
Validation accuracy = 82.03%
Test accuracy = 76.6 %

Pruning strategy:
For each node I have stored the index of validation data passing through it. In each phase of pruning, the algorithm goes to all the nodes which are parent of leaf nodes only. It checks whether pruning it increases the accuracy (on that stored indexes data only) or not. If accuracy decreases then it doesn't otherwise it prunes it.

(1.C)



Training accuracy = 0.997889
Validation accuracy = 0.718000
Test accuracy = 0.721500
Preprocessed data accuracy of part(A) is better because overfitting happens here.

(1.D)

Parameters:

- min_samples_split = 9
- min_samples_leaf = 4
- max_depth = 10
- random_state = 0

Training accuracy = 85.44 %
Validation accuracy = 79.96 %

Test accuracy = 80.15 %
Validation accuracy decreases as max_depth is increased from 10. Below this depth the majority values start to dominate and accuracy approaches majority accuracy and f_score reduces.
Compared to part b and c above, the accuracy is better in this case if we look at f_score as this tree is good in predicting the minority class too.

(1.E)
Train accuracy = 85.09%
Validation accuracy = 79.60%
Test accuracy = 80.42 %
Looks like there is not much change even after one hot encoding. Maybe because this is the best range of accuracy these trees can reach also the dataset is skewed (77% majority class) maybe it depends on that too. Let's look at the random forest results

(1.F)
Parameters:

- n_estimators=100
- max_depth=10
- random_state=0
- bootstrap=False

Accuracy increases with n_estimators till it saturates at a certain value.
Gini criterion seems to work better than entropy
Bootstrap = False seems to give a slightly better accuracy.
Train accuracy = 87.21%
Validation accuracy = 80.32%
Test accuracy = 80.83%
As guessed, the accuracy of random forests too seem to stabilize around this range.

Question 2:

(2.A)

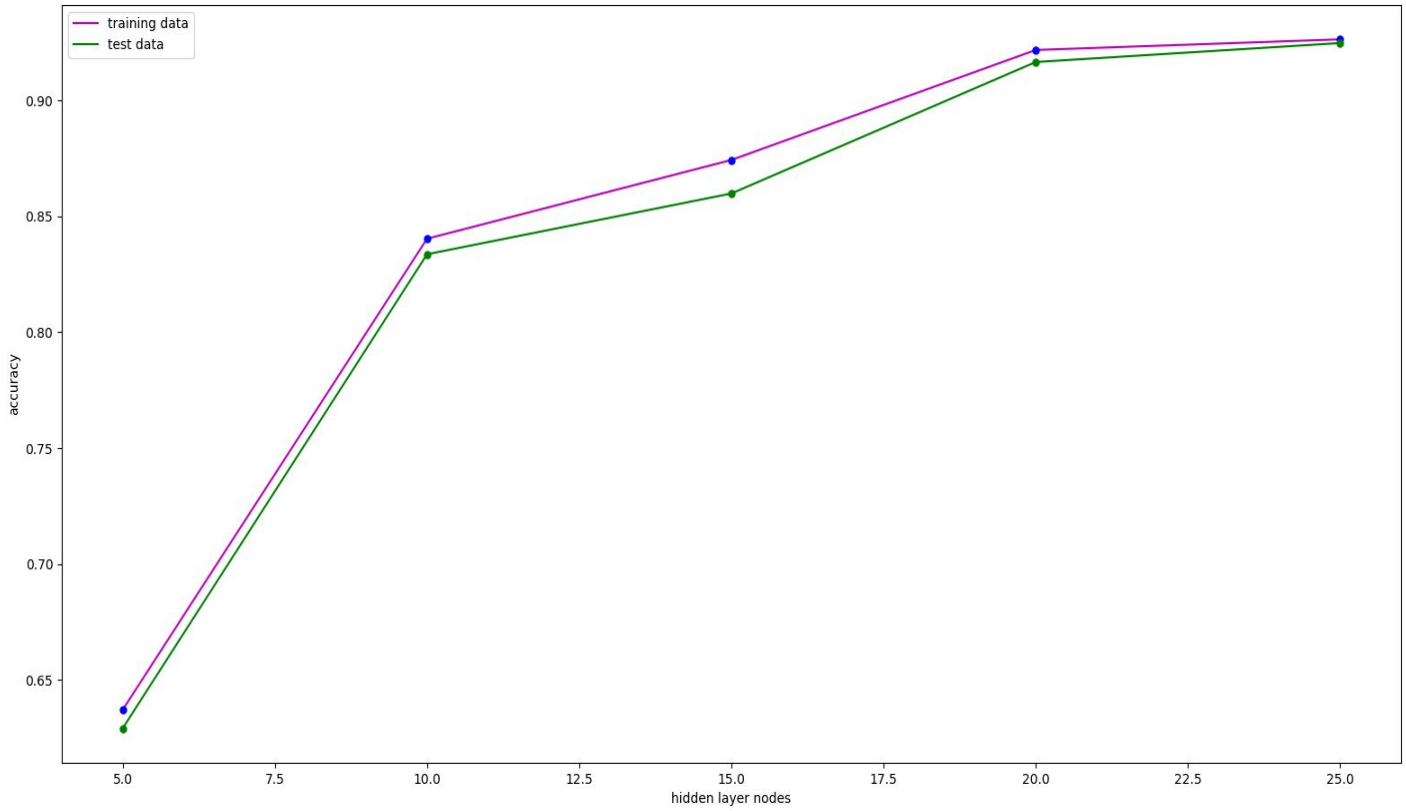
One hot encoded data drive link:
https://drive.google.com/drive/folders/1Gn-p6wvxqYwoZ3gPqOhMa1zJVTGQNtap?usp=sharing

(2.B)

Coding part

(2.C)

For one hidden layer



Training time(in seconds) = 52,241,86,26,28 respectively
"error" is calculated on validation set.

Stopping criteria:
(If after 80 epochs the current error is greater than 80 epochs before) or (the absolute change in error after 80 epochs is less than 10^{-6}) or (total runtime>10 mins)

Model stores weights and bias which gives least error on validation set over the whole runtime

Confusion matrices:

```
1 hidden layer 5 nodes:
training time = 52 seconds
train accuracy = 63.730508 percent
test accuracy = 62.903800 percent
[[375602 125607      0      0      0      0      0      0      0      0]
 [169062 253436      0      0      0      0      0      0      0      0]
 [  7871  39751      0      0      0      0      0      0      0      0]
 [  6515  14606      0      0      0      0      0      0      0      0]
 [  1674   2211      0      0      0      0      0      0      0      0]
 [  1487    509      0      0      0      0      0      0      0      0]
 [   126   1298      0      0      0      0      0      0      0      0]
 [    24    206      0      0      0      0      0      0      0      0]
 [     6      6      0      0      0      0      0      0      0      0]
 [     0      3      0      0      0      0      0      0      0      0]]
```

```
1 hidden layer 10 nodes:
training time = 241 seconds
train accuracy = 84.042383 percent
test accuracy = 83.379800 percent
[[485452  15757      0      0      0      0      0      0      0      0]
 [ 74152 348346      0      0      0      0      0      0      0      0]
 [   354  47268      0      0      0      0      0      0      0      0]
 [  1511  19610      0      0      0      0      0      0      0      0]
 [  2861   1024      0      0      0      0      0      0      0      0]
 [  1930     66      0      0      0      0      0      0      0      0]
 [     1   1423      0      0      0      0      0      0      0      0]
 [     1    229      0      0      0      0      0      0      0      0]
 [     6      6      0      0      0      0      0      0      0      0]
 [     0      3      0      0      0      0      0      0      0      0]]
```

```
1 hidden layer 15 nodes:
training time = 86 seconds
train accuracy = 87.445022 percent
test accuracy = 85.993600 percent
[[481773  19436      0      0      0      0      0      0      0      0]
 [ 44665 377828      1      0      2      0      0      0      0      2]
 [    17  47270    335      0      0      0      0      0      0      0]
 [   265  20769     83      0      4      0      0      0      0      0]
 [  3584    301      0      0      0      0      0      0      0      0]
 [  1915     81      0      0      0      0      0      0      0      0]
 [     0   1145    279      0      0      0      0      0      0      0]
 [     0    190     40      0      0      0      0      0      0      0]
 [    10      2      0      0      0      0      0      0      0      0]
 [     3      0      0      0      0      0      0      0      0      0]]
```

```
1 hidden layer 20 nodes:
training time = 26 seconds
train accuracy = 92.191124 percent
test accuracy = 91.674300 percent
[[498013   3196      0      0      0      0      0      0      0      0]
 [  6920 415548      0     30      0      0      0      0      0      0]
 [     1  45835   1769     17      0      0      0      0      0      0]
 [   141  19289    268   1413      0      0      0      0      0     10]
 [  3660    225      0      0      0      0      0      0      0      0]
 [  1972     24      0      0      0      0      0      0      0      0]
 [     0    665    697     62      0      0      0      0      0      0]
 [     1    131     78     20      0      0      0      0      0      0]
 [    12      0      0      0      0      0      0      0      0      0]
 [     2      1      0      0      0      0      0      0      0      0]]
```

```
1 hidden layer 25 nodes:
training time = 28 seconds
train accuracy = 92.658936 percent
test accuracy = 92.491200 percent
[[499162   2047      0      0      0      0      0      0      0      0]
 [  3758 418662      5     73      0      0      0      0      0      0]
 [     0  40583   7039      0      0      0      0      0      0      0]
 [   161  20551    360     49      0      0      0      0      0      0]
 [  3712    173      0      0      0      0      0      0      0      0]
 [  1960     36      0      0      0      0      0      0      0      0]
 [     0    658    766      0      0      0      0      0      0      0]
 [     7    147     71      5      0      0      0      0      0      0]
 [    11      1      0      0      0      0      0      0      0      0]
 [     3      0      0      0      0      0      0      0      0      0]]
```

In 1 hidden layer case, as the number of nodes increases the number of true positives increases also the accuracy.

(2.D)

```
training time = 28 seconds
train accuracy = 67.061176 percent
test accuracy = 65.814400 percent
[[459629  41580      0      0      0      0      0      0      0      0]
 [223983 198515      0      0      0      0      0      0      0      0]
 [  9734  37888      0      0      0      0      0      0      0      0]
 [  8566  12555      0      0      0      0      0      0      0      0]
 [  3303    582      0      0      0      0      0      0      0      0]
 [  1835    161      0      0      0      0      0      0      0      0]
 [    55   1369      0      0      0      0      0      0      0      0]
 [    52    178      0      0      0      0      0      0      0      0]
 [    11      1      0      0      0      0      0      0      0      0]
 [     3      0      0      0      0      0      0      0      0      0]]
```

```
2 hidden layer with 10 nodes each:

training time = 53 seconds
train accuracy = 78.648541 percent
test accuracy = 77.166300 percent
[[438977  62211      1      0     20      0      0      0      0      0]
 [ 97883 319400   5209      0      6      0      0      0      0      0]
 [   669  33687  13266      0      0      0      0      0      0      0]
 [  2288  16039   2794      0      0      0      0      0      0      0]
 [  3512    353      0      0     20      0      0      0      0      0]
 [  1724    272      0      0      0      0      0      0      0      0]
 [     3    823    598      0      0      0      0      0      0      0]
 [     4    162     64      0      0      0      0      0      0      0]
 [    12      0      0      0      0      0      0      0      0      0]
 [     2      1      0      0      0      0      0      0      0      0]]
```

```
2 hidden layer with 15 nodes each:

training time = 59 seconds
train accuracy = 97.816873 percent
test accuracy = 97.075500 percent
[[501095    111      0      3      0      0      0      0      0      0]
 [  2250 417067    689   2492      0      0      0      0      0      0]
 [     8   4251  41535   1827      0      0      1      0      0      0]
 [   417   8724    922  11057      0      0      1      0      0      0]
 [  3874     11      0      0      0      0      0      0      0      0]
 [  1996      0      0      0      0      0      0      0      0      0]
 [     1    242   1125     55      0      0      1      0      0      0]
 [     3     59      6    162      0      0      0      0      0      0]
 [    12      0      0      0      0      0      0      0      0      0]
 [     3      0      0      0      0      0      0      0      0      0]]
```

```
2 hidden layers with 20 nodes each:

training time = 52 seconds
train accuracy = 99.132347 percent
test accuracy = 98.316400 percent
[[499521    294      0      0     83   1279      0      0      0     32]
 [   461 421136    322    564      4      7      0      0      0      4]
 [     0   1106  44336   2143      0      0     37      0      0      0]
 [     1   2042   1048  18027      0      0      3      0      0      0]
 [  3834     26      0      0     13     12      0      0      0      0]
 [  1902      3      0      0      4     85      0      0      0      2]
 [     0     13   1280     85      0      0     46      0      0      0]
 [     0      6     37    187      0      0      0      0      0      0]
 [     9      1      0      0      1      1      0      0      0      0]
 [     1      0      0      0      0      2      0      0      0      0]]
```

```
2 hidden layers with 25 nodes each:

training time = 29 seconds
train accuracy = 99.796082 percent
test accuracy = 99.163700 percent
[[499963      2      0      0    972    264      0      0      0      8]
 [     0 422496      0      2      0      0      0      0      0      0]
 [     0      0  47581      1      0      0     40      0      0      0]
 [     0    290    174  20657      0      0      0      0      0      0]
 [  3324      2      0      0    548      7      0      0      0      4]
 [  1604      0      0      0     15    375      0      0      0      2]
 [     0      0   1407      0      0      0     17      0      0      0]
 [     0      0     16    214      0      0      0      0      0      0]
 [     6      0      0      0      2      4      0      0      0      0]
 [     1      0      0      0      0      2      0      0      0      0]]
```

Keeping the number of units in hidden layers same, if we increase the number of hidden layers then f score increases (diagonal elements increases) along with accuracy.

(2.E)

Variable learning rate. I have used tolerance = 10^{-6}

```
1 hidden layer 5 nodes:
training time = 20 seconds
train accuracy = 64.070372 percent
test accuracy = 63.269500 percent
[[415053  86156      0      0      0      0      0      0      0      0]
 [204856 217642      0      0      0      0      0      0      0      0]
 [ 11362  36260      0      0      0      0      0      0      0      0]
 [  8376  12745      0      0      0      0      0      0      0      0]
 [  2203   1682      0      0      0      0      0      0      0      0]
 [  1652    344      0      0      0      0      0      0      0      0]
 [   203   1221      0      0      0      0      0      0      0      0]
 [    35    195      0      0      0      0      0      0      0      0]
 [     8      4      0      0      0      0      0      0      0      0]
 [     0      3      0      0      0      0      0      0      0      0]]
```

```
1 hidden layer 10 nodes:
training time = 63 seconds
train accuracy = 85.349860 percent
test accuracy = 84.670500 percent
[[486843  14366      0      0      0      0      0      0      0      0]
 [ 62636 359862      0      0      0      0      0      0      0      0]
 [   227  47395      0      0      0      0      0      0      0      0]
 [  1153  19968      0      0      0      0      0      0      0      0]
 [  3367    518      0      0      0      0      0      0      0      0]
 [  1944     52      0      0      0      0      0      0      0      0]
 [     4   1420      0      0      0      0      0      0      0      0]
 [     0    230      0      0      0      0      0      0      0      0]
 [     9      3      0      0      0      0      0      0      0      0]
 [     3      0      0      0      0      0      0      0      0      0]]
```

```
1 hidden layer 15 nodes:
training time = 46 seconds
train accuracy = 88.160736 percent
test accuracy = 86.463100 percent
[[483841  17368      0      0      0      0      0      0      0      0]
 [ 41899 380595      4      0      0      0      0      0      0      0]
 [     8  47419    195      0      0      0      0      0      0      0]
 [   174  20915     32      0      0      0      0      0      0      0]
 [  3501    384      0      0      0      0      0      0      0      0]
 [  1931     65      0      0      0      0      0      0      0      0]
 [     0   1193    231      0      0      0      0      0      0      0]
 [     1    192     36      0      1      0      0      0      0      0]
 [     9      3      0      0      0      0      0      0      0      0]
 [     3      0      0      0      0      0      0      0      0      0]]
```

```
1 hidden layer 20 nodes:
training time = 40 seconds
train accuracy = 92.475010 percent
test accuracy = 92.077300 percent
[[499595   1614      0      0      0      0      0      0      0      0]
 [  3015 419483      0      0      0      0      0      0      0      0]
 [     1  47278    342      1      0      0      0      0      0      0]
 [   157  19520     91   1353      0      0      0      0      0      0]
 [  3768    117      0      0      0      0      0      0      0      0]
 [  1982     14      0      0      0      0      0      0      0      0]
 [     0    751    602     71      0      0      0      0      0      0]
 [     3    131     78     18      0      0      0      0      0      0]
 [    12      0      0      0      0      0      0      0      0      0]
 [     2      1      0      0      0      0      0      0      0      0]]
```

```
1 hidden layer 25 nodes:
training time = 27 seconds
train accuracy = 92.319072 percent
test accuracy = 92.158700 percent
[[499938   1271      0      0      0      0      0      0      0      0]
 [  2128 420370      0      0      0      0      0      0      0      0]
 [     6  46337   1279      0      0      0      0      0      0      0]
 [    26  21004     91      0      0      0      0      0      0      0]
 [  3775    110      0      0      0      0      0      0      0      0]
 [  1982     14      0      0      0      0      0      0      0      0]
 [     0    714    710      0      0      0      0      0      0      0]
 [     0    124    106      0      0      0      0      0      0      0]
 [    12      0      0      0      0      0      0      0      0      0]
 [     3      0      0      0      0      0      0      0      0      0]]
```

```
2 hidden layers 5 nodes each:
training time = 36 seconds
train accuracy = 67.552979 percent
test accuracy = 65.800200 percent
[[432650  68559      0      0      0      0      0      0      0      0]
 [197146 225352      0      0      0      0      0      0      0      0]
 [  8209  39413      0      0      0      0      0      0      0      0]
 [  6668  14453      0      0      0      0      0      0      0      0]
 [  2535   1350      0      0      0      0      0      0      0      0]
 [  1741    255      0      0      0      0      0      0      0      0]
 [    38   1386      0      0      0      0      0      0      0      0]
 [    31    199      0      0      0      0      0      0      0      0]
 [    10      2      0      0      0      0      0      0      0      0]
 [     3      0      0      0      0      0      0      0      0      0]]
```

```
2 hidden layers 10 nodes each:
training time = 69 seconds
train accuracy = 80.927629 percent
test accuracy = 78.984200 percent
[[441138  59943    128      0      0      0      0      0      0      0]
 [ 84556 331430   6512      0      0      0      0      0      0      0]
 [  1209  29139  17274      0      0      0      0      0      0      0]
 [  1173  17576   2372      0      0      0      0      0      0      0]
 [  3560    325      0      0      0      0      0      0      0      0]
 [  1763    232      1      0      0      0      0      0      0      0]
 [     4    714    706      0      0      0      0      0      0      0]
 [     0    158     72      0      0      0      0      0      0      0]
 [    11      1      0      0      0      0      0      0      0      0]
 [     3      0      0      0      0      0      0      0      0      0]]
```

```
2 hidden layers 15 nodes each:
training time = 33 seconds
train accuracy = 96.797281 percent
test accuracy = 96.304500 percent
[[500985    224      0      0      0      0      0      0      0      0]
 [  1269 420429    800      0      0      0      0      0      0      0]
 [     8   5983  41631      0      0      0      0      0      0      0]
 [   141  14386   6594      0      0      0      0      0      0      0]
 [  3646    239      0      0      0      0      0      0      0      0]
 [  1994      2      0      0      0      0      0      0      0      0]
 [     0    240   1184      0      0      0      0      0      0      0]
 [     0    108    122      0      0      0      0      0      0      0]
 [    10      2      0      0      0      0      0      0      0      0]
 [     3      0      0      0      0      0      0      0      0      0]]
```

```
2 hidden layers 20 nodes each:
training time = 24 seconds
train accuracy = 97.349060 percent
test accuracy = 96.938800 percent
[[501099    110      0      0      0      0      0      0      0      0]
 [   137 422120    197     44      0      0      0      0      0      0]
 [     0   3266  43881    475      0      0      0      0      0      0]
 [     0  10398   8435   2288      0      0      0      0      0      0]
 [  3884      1      0      0      0      0      0      0      0      0]
 [  1995      1      0      0      0      0      0      0      0      0]
 [     0     12   1410      2      0      0      0      0      0      0]
 [     0     63    148     19      0      0      0      0      0      0]
 [    12      0      0      0      0      0      0      0      0      0]
 [     3      0      0      0      0      0      0      0      0      0]]
```

```
2 hidden layers 25 nodes each:
training time = 65 seconds
train accuracy = 99.704118 percent
test accuracy = 99.154800 percent
[[500090      0      0      0    838    281      0      0      0      0]
 [     0 422498      0      0      0      0      0      0      0      0]
 [     0      0  47568     54      0      0      0      0      0      0]
 [     0    575     30  20516      0      0      0      0      0      0]
 [  3338      1      0      0    528     18      0      0      0      0]
 [  1603      0      0      0     45    348      0      0      0      0]
 [     0      0   1424      0      0      0      0      0      0      0]
 [     0      1      1    228      0      0      0      0      0      0]
 [     5      0      0      0      4      3      0      0      0      0]
 [     1      0      0      0      0      2      0      0      0      0]]
```

(2.F)

ReLU seems to achieve at max of 50% accuracy. Looks like there is some problem in the implementation of ReLU method even though I have implemented it.