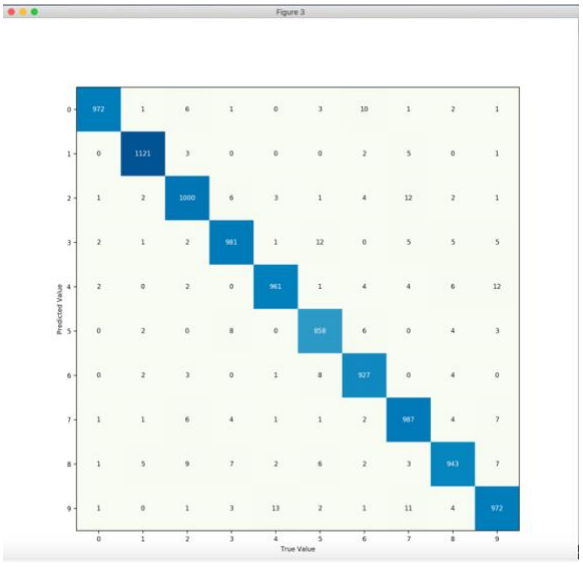
Project Report

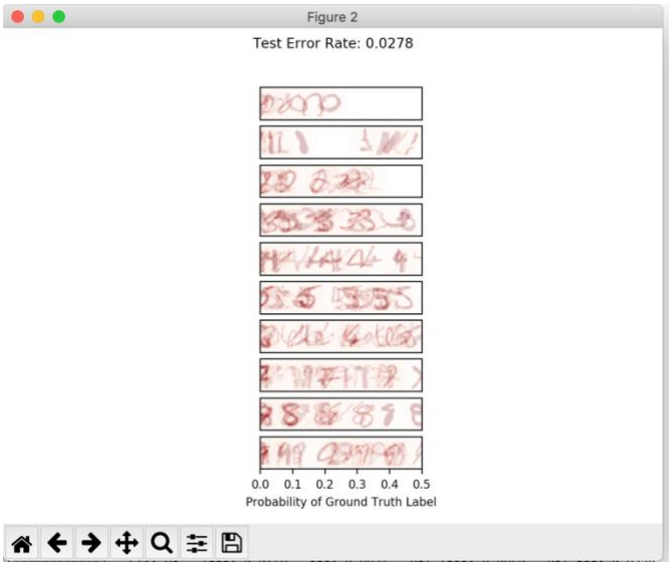
Most optimal Values :

Epoch Value = 32
Batch size = 32
alpha Value = 0.1
Test error rate = 0.0290
Test accuracy = 0.9722
Train accuracy = 0. 9988
Train loss = 0.0087
Validation accuracy = 0.9742
Validation loss = 0.0942

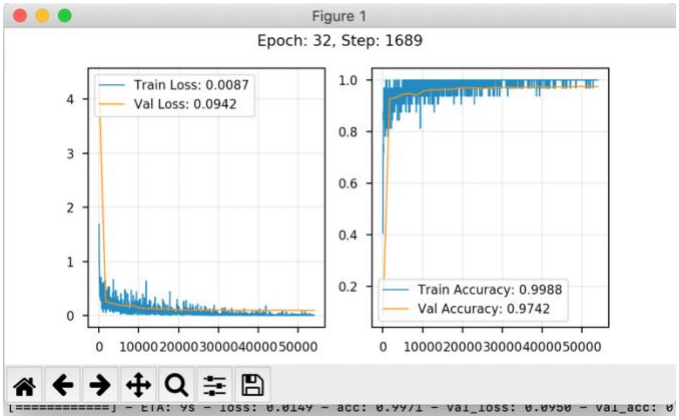
Hidden layers size value I have used are 256 and 256 and I have used sigmoid functions to activate them.

I intitialized all the bias (b1,b2,b3) to 0
I have initialized the weights (w1,w2,w3) to a small random value near zero (according to a zero-meaned normal distribution) with range (-0.5 to 0.5).





```
 54016/54000 [==============] - ETA: 10s - loss: 0.1059 - acc: 0.9680 - val_loss: 0.1314 - val_acc: 0.9612
Epoch: 8/32
 54016/54000 [==============] - ETA: 9s - loss: 0.0935 - acc: 0.9719 - val_loss: 0.1276 - val_acc: 0.9620
Epoch: 9/32
 54016/54000 [==============] - ETA: 9s - loss: 0.0836 - acc: 0.9749 - val_loss: 0.1193 - val_acc: 0.9645
Epoch: 10/32
 54016/54000 [==============] - ETA: 9s - loss: 0.0748 - acc: 0.9774 - val_loss: 0.1198 - val_acc: 0.9632
Epoch: 11/32
 54016/54000 [==============] - ETA: 9s - loss: 0.0674 - acc: 0.9799 - val_loss: 0.1132 - val_acc: 0.9675
Epoch: 12/32
 54016/54000 [==============] - ETA: 9s - loss: 0.0599 - acc: 0.9824 - val_loss: 0.1037 - val_acc: 0.9702
Epoch: 13/32
 54016/54000 [==============] - ETA: 9s - loss: 0.0545 - acc: 0.9841 - val_loss: 0.1025 - val_acc: 0.9695
Epoch: 14/32
 54016/54000 [==============] - ETA: 9s - loss: 0.0490 - acc: 0.9857 - val_loss: 0.1015 - val_acc: 0.9700
Epoch: 15/32
 54016/54000 [==============] - ETA: 9s - loss: 0.0447 - acc: 0.9875 - val_loss: 0.1082 - val_acc: 0.9670
Epoch: 16/32
 54016/54000 [==============] - ETA: 9s - loss: 0.0407 - acc: 0.9889 - val_loss: 0.1000 - val_acc: 0.9693
Epoch: 17/32
 54016/54000 [==============] - ETA: 9s - loss: 0.0364 - acc: 0.9905 - val_loss: 0.1025 - val_acc: 0.9708
Epoch: 18/32
 54016/54000 [==============] - ETA: 9s - loss: 0.0328 - acc: 0.9916 - val_loss: 0.0974 - val_acc: 0.9713
Epoch: 19/32
 54016/54000 [==============] - ETA: 9s - loss: 0.0300 - acc: 0.9924 - val_loss: 0.0986 - val_acc: 0.9717
Epoch: 20/32
 54016/54000 [==============] - ETA: 9s - loss: 0.0271 - acc: 0.9930 - val_loss: 0.0978 - val_acc: 0.9707
Epoch: 21/32
 54016/54000 [==============] - ETA: 9s - loss: 0.0242 - acc: 0.9943 - val_loss: 0.0948 - val_acc: 0.9730
Epoch: 22/32
 54016/54000 [==============] - ETA: 9s - loss: 0.0220 - acc: 0.9950 - val_loss: 0.0959 - val_acc: 0.9730
Epoch: 23/32
 54016/54000 [==============] - ETA: 9s - loss: 0.0199 - acc: 0.9954 - val_loss: 0.0919 - val_acc: 0.9747
Epoch: 24/32
 54016/54000 [==============] - ETA: 9s - loss: 0.0183 - acc: 0.9962 - val_loss: 0.0972 - val_acc: 0.9705
Epoch: 25/32
 54016/54000 [==============] - ETA: 9s - loss: 0.0167 - acc: 0.9964 - val_loss: 0.0929 - val_acc: 0.9742
Epoch: 26/32
 54016/54000 [==============] - ETA: 9s - loss: 0.0149 - acc: 0.9971 - val_loss: 0.0950 - val_acc: 0.9738
Epoch: 27/32
 54016/54000 [==============] - ETA: 9s - loss: 0.0136 - acc: 0.9977 - val_loss: 0.0984 - val_acc: 0.9718
Epoch: 28/32
 54016/54000 [==============] - ETA: 9s - loss: 0.0126 - acc: 0.9979 - val_loss: 0.0942 - val_acc: 0.9738
Epoch: 29/32
 54016/54000 [==============] - ETA: 9s - loss: 0.0113 - acc: 0.9983 - val_loss: 0.0949 - val_acc: 0.9745
Epoch: 30/32
 54016/54000 [==============] - ETA: 9s - loss: 0.0101 - acc: 0.9985 - val_loss: 0.0949 - val_acc: 0.9750
Epoch: 31/32
 54016/54000 [==============] - ETA: 9s - loss: 0.0095 - acc: 0.9986 - val_loss: 0.0936 - val_acc: 0.9755
Epoch: 32/32
 54016/54000 [==============] - ETA: 9s - loss: 0.0087 - acc: 0.9988 - val_loss: 0.0942 - val_acc: 0.9742
Test...
Test Accuracy: 0.9722
Learning Rate
0.1
Batch Size
32
Max Iterations
32
Hidden layers -
256
256
(base) Kartiks-MacBook-Pro:project5 kartikrao$
```



I then tried to use the same architecture on the alphabet (a-i) (not_mnist) datasets which gave me a -

Test error rate = 0.0756
Test accuracy = 0.9244
Train accuracy = 0. 9900
Train loss = 0.0600
Validation accuracy = 0.8605
Validation loss = 0.6265