CS440 MP 3

Part 1: Digit classification

Digit classification part of MP3 was implemented with Python 3.

Reading the Data

In our implementation, each entry from **training** or **test data** are converted to 28x28 matrix of 1s and 0s. Empty spaces are converted to 0 and every other element is converted to 1.

Let $\boldsymbol{Train_i}$ be a 28x28 matrix (with entries 0 and 1) representing an entry from the training data, where $i = 1,..,5000$.

Let $\boldsymbol{Test_i}$ be a 28x28 matrix (with entries 0 and 1) representing an entry from the training data, where $i = 1,..,1000$.

Training

$Prob(\ F_{ab} = 1 \mid class = digit\ x)$, $Prob(\ F_{ab} = 0 \mid class = digit\ x)$, and $Prob(class = digit\ i)$ are computed in equation 1 ,2 and 3, respectively.

Let $\boldsymbol{P_x}$ be a 28x28 matrix where $P_x[a][b] = Prob(\ F_{ab} = 1 \mid class = digit\ x)$

[1] $$\boldsymbol{P_x} = \left(\sum_{i\ is\ digit\ x} Train_i + ones(28,28) * k\right) \frac{1}{(5000)+2*k}$$

Where the **Number of training data = 5000** and $\boldsymbol{k} \equiv \boldsymbol{Laplacian\ smoothing\ factor}$. Ones(28,28) is a 28x28 matrix where all the entries are 1.

[2] $$\boldsymbol{Prob(\ F_{ab} = 0 \mid class = digit\ x)} = 1 - Prob(\ F_{ab} = 1 \mid class = digit\ x)$$

[3] $$Prob(class = digit\ i) = \frac{\#\ of\ digit\ i\ in\ training\ data}{5000}$$

Identifying the image

$Test_i$ is identified as digit z:

[4] $$z = \text{argmax}\{\ (Likelihood\ Test_i = digit\ 1)..,(Likelihood\ Test_i = digit\ 10)\ \}$$

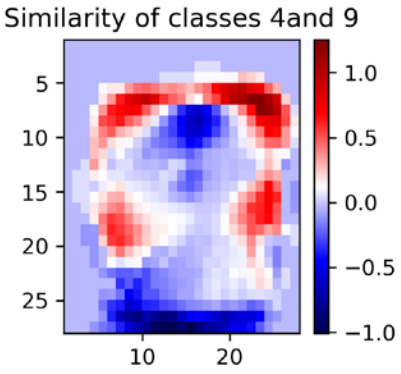
[5] (Likelihood $Test_i$ =digit x)= $\log[\ Prob(class = digit\ x)] + \sum_{a=1,b=1}^{28,28} \log[\ Prob(F_{a,b} \mid class = digit\ x)]$

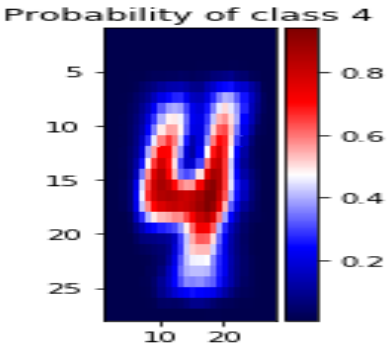
Results:

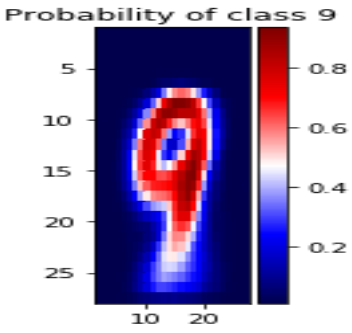
Laplacian smoothing factor, **k =5**, was used.

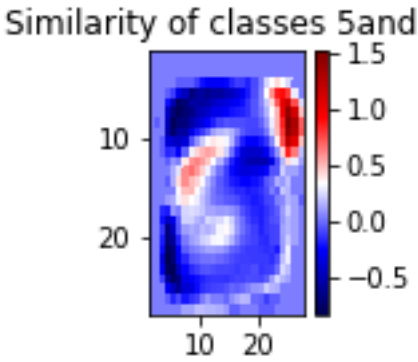
```
-----------Correct classification rate--------------
[84.44, 96.29, 73.78, 79.0, 75.700, 71.21, 71.43, 71.69, 70.194174757281552, 79.0]
```

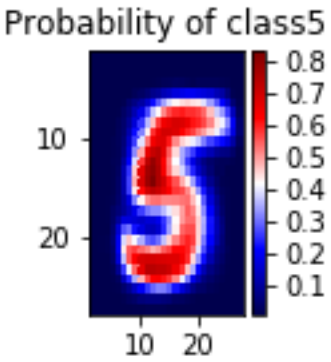
```
---------- Confusion Matrix-------------
[[ 84.44444444   0.           1.11111111   0.           1.11111111   5.55555556   4.44444444   0.           3.33333333   0.        ]
 [  0.          96.2962963    0.92592593   0.           0.           1.85185185   0.92592593   0.           0.           0.        ]
 [  1.94174757   3.88349515  73.78640777   5.82524272   0.97087379   0.           5.82524272   0.97087379   4.85436893   1.94174757]
 [  0.           2.           0.          79.           0.           3.           2.           6.           2.           6.        ]
 [  0.           0.93457944   0.           0.          75.70093458   0.           2.80373832   0.93457944   1.86915888  17.75700935]
 [  2.17391304   2.17391304   1.08695652  15.2173913    3.26086957  71.2123913    1.08695652   1.08695652   2.17391304   6.52173913]
 [  1.0989011    6.59340659   5.49450549   0.           7.69230769   5.49450549  71.42857143   0.           2.1978022    0.        ]
 [  0.           6.60377358   2.83018868   0.           2.83018868   0.           0.          71.69811321   1.88679245  14.1509434 ]
 [  1.94174757   1.94174757   2.91262136  12.62135922   1.94174757   4.85436893   0.97087379   0.97087379  70.19417476  11.65048544]
 [  1.           2.           1.           3.           9.           2.           0.           2.           1.          79.        ]]
```

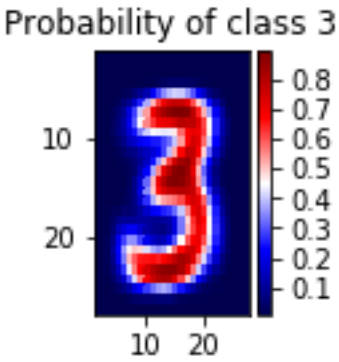


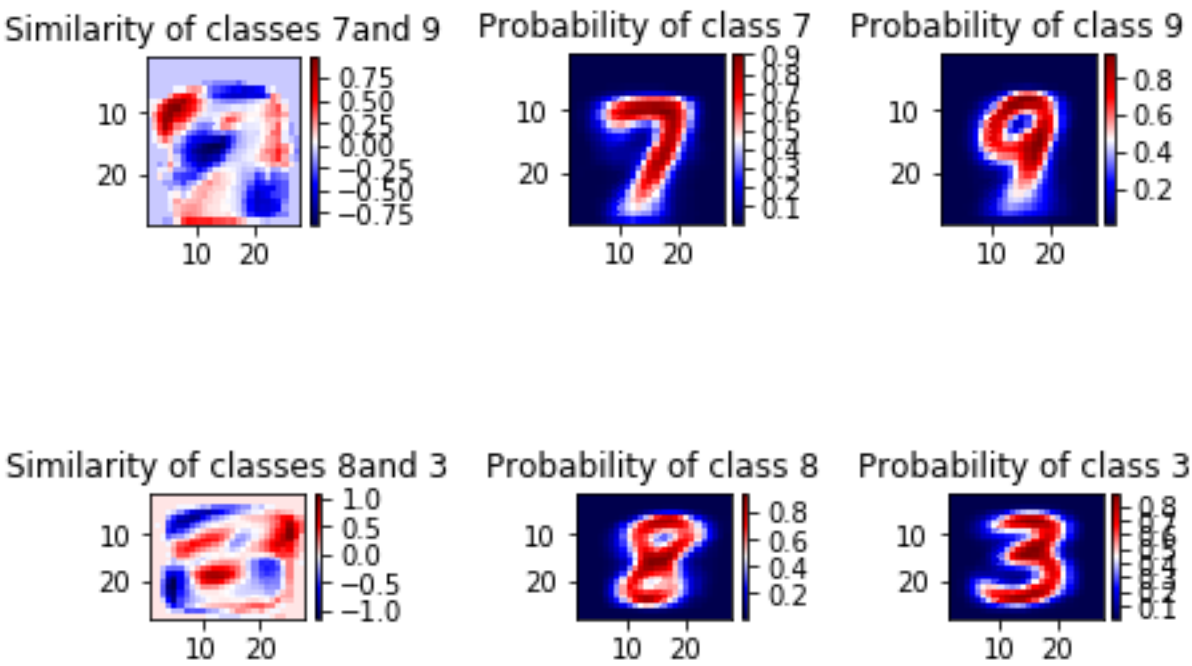











Similarity of classes 7and 9
Probability of class 7
Probability of class 9
Similarity of classes 8and 3
Probability of class 8
Probability of class 3