HOMEWORK 7

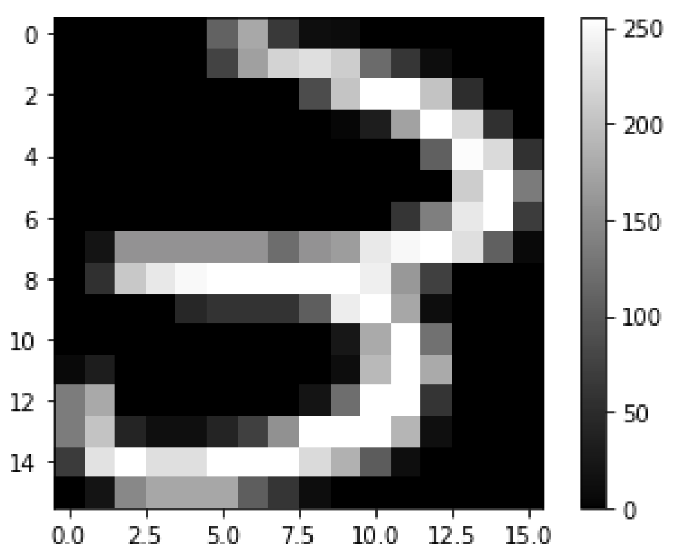
Hafeez Ali Anees Ali
9084603423

1 Principal Component Analysis [60 pts]

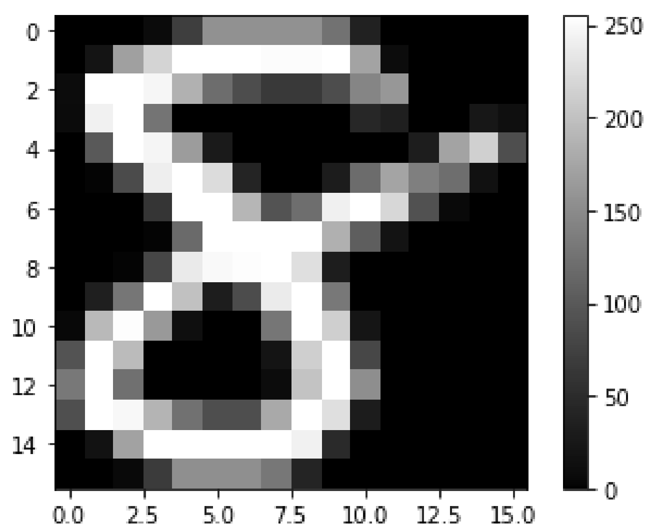
Download three.txt and eight.txt. Each has 200 handwritten digits. Each line is for a digit, vectorized from a 16 × 16 grayscale image.

1. (10 pts) Each line has 256 numbers: They are pixel values (0=black, 255=white) vectorized from the image as the first column (top-down), the second column, and so on. Visualize the two grayscale images corresponding to the first line in three.txt and the first line in eight.txt.

Sol:



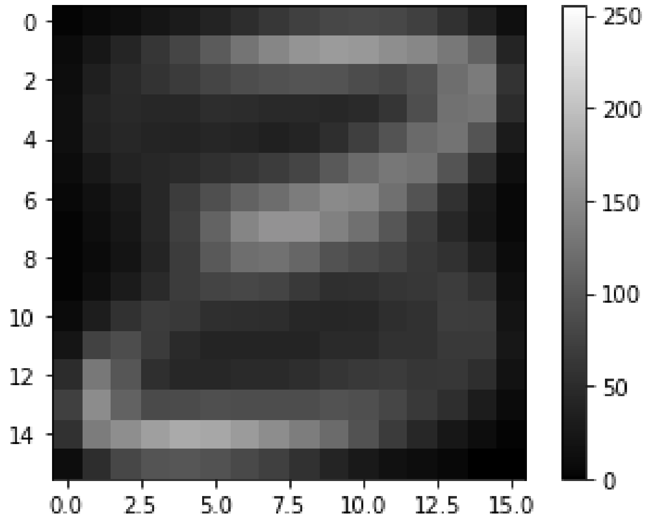
First line in three.txt



First line in eight.txt

2. (10 pts) Putting the two data files together (three first, eight next) to form an $n \times D$ matrix X where $n = 400$ digits and $D = 256$ pixels. Note we use $n \times D$ size for X instead of $D \times n$ to be consistent with the convention in linear regression. The i-th row of X is $x^{\top}_i$, where $x_i \in R^D$ is the i-th image in the combined data set. Compute the sample mean y. Visualize y as a 16×16 grayscale image.

Sol:



Sample mean Y

3. (10 pts) Center X using y above. Then form the sample covariance matrix S and show the 5x5 submatrix S[0:5, 0:5]

Sol:

```
S_submatrix_5x5
```

```
array([[  59.16729323,  142.14943609,   28.68201754,   -7.17857143,
         -14.3358396 ],
       [ 142.14943609,  878.93879073,  374.13731203,   24.12778195,
         -87.12781955],
       [  28.68201754,  374.13731203, 1082.9058584 ,  555.2268797 ,
          33.72431078],
       [  -7.17857143,   24.12778195,  555.2268797 , 1181.24408521,
         777.77192982],
       [ -14.3358396 ,  -87.12781955,   33.72431078,  777.77192982,
        1429.95989975]])
```

5 x 5 submatrix S(1…5, 1…5)

4. (10 pts) Use appropriate software to compute the two largest eigenvalues $\lambda_1 \geq \lambda_2$ and the corresponding eigenvectors v^1, v^2 of S. For example, in Matlab one can use eigs(S,2). Show the value of λ_1, λ_2. Visualize v^1, v^2as two 16×16 grayscale images. Hint: Their elements will not be in [0, 255], but you can shift and scale them appropriately. It is best if you can show an accompanying 'colorbar' that maps the grayscale to values.

Sol:

Eigenvalues L1 and L2 are: **237155.24** & **145188.35** respectively.

```
L1
```

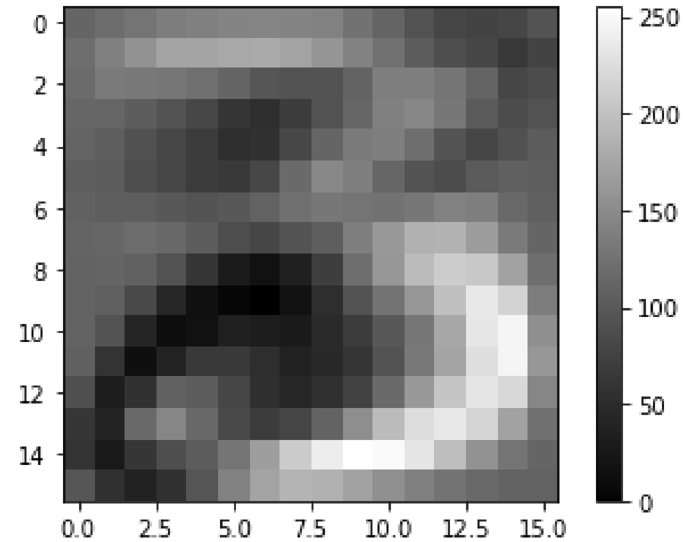
237155.2462904853

```
L2
```

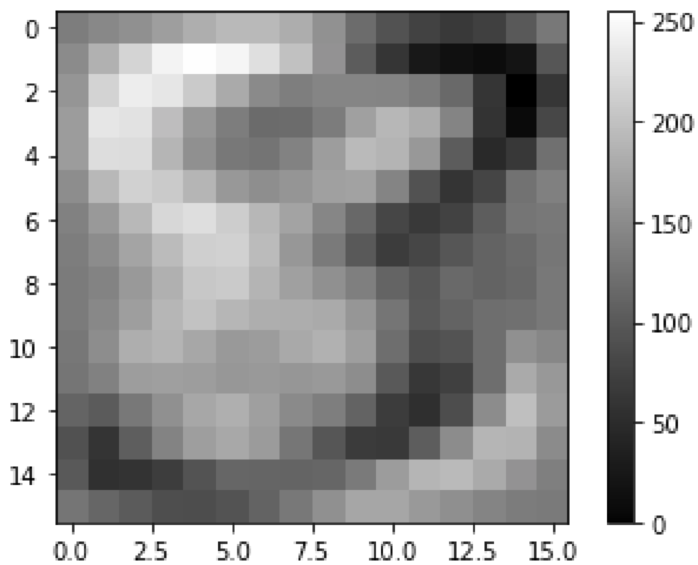
145188.3526868256

Eigenvectors v1 and v2 visualized as grayscale images:

V1:



V2:



For the image dataset of threes and eights, the PCA algorithm identifies the above two images as orthogonal components (eigenvectors). As can be seen from the images, for V1, the pixels that would be part of 3 are closer to 0 and pixels that are part of 8 are closer to 255. For V2, the pixels part of 3 are more activated than 8. These two eigenvectors help provide maximum data reconstruction for two dimensions. The combination of these two eigenvectors also help differentiate threes and eights in two dimensions (linearly separable with soft margin) as can be seen in Q6.

5. (10 pts) Now we project (the centered) X down to the two PCA directions. Let V = [v^1 v^2] be the $D \times 2$ matrix. The projection is simply XV. Show the resulting two coordinates for the first line in three.txt and the first line in eight.txt, respectively.

Sol:

Coordinates for the first line in three.txt: (228352.83, -49638.115)

Coordinates for the first line in eight.txt: (251859.83, 1135966.885)

```
# co-ordinates for 1st three
XV_image[0]
```

```
array([228352.83 , -49638.115])
```

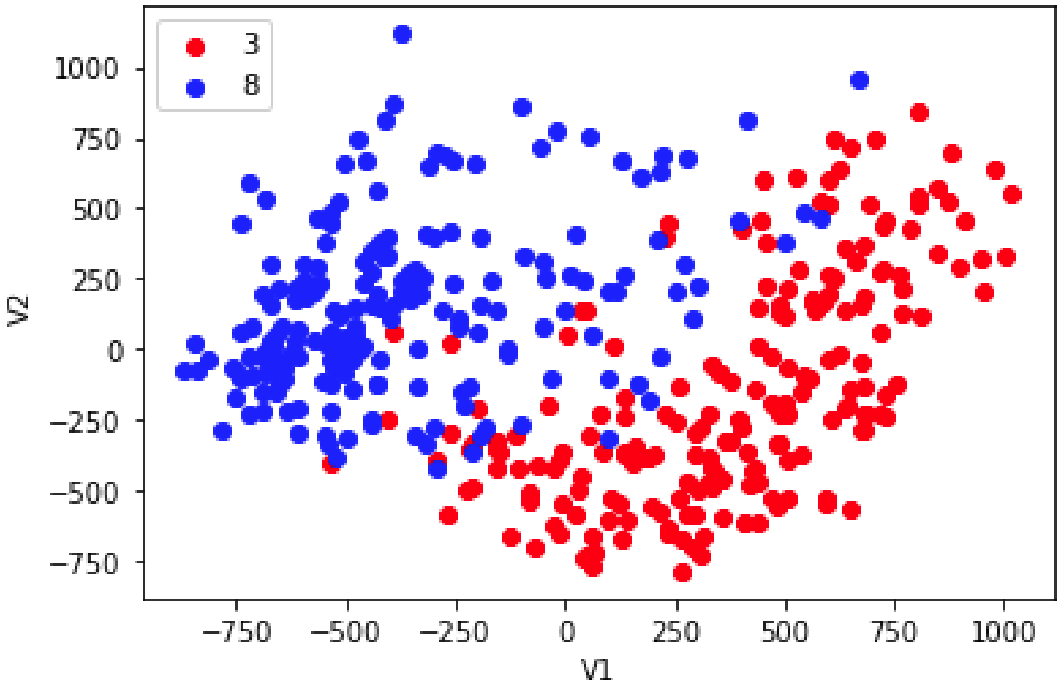
```
# co-ordinates for 1st eight
XV_image[200]
```

```
array([ 251859.83 , 1135966.885])
```

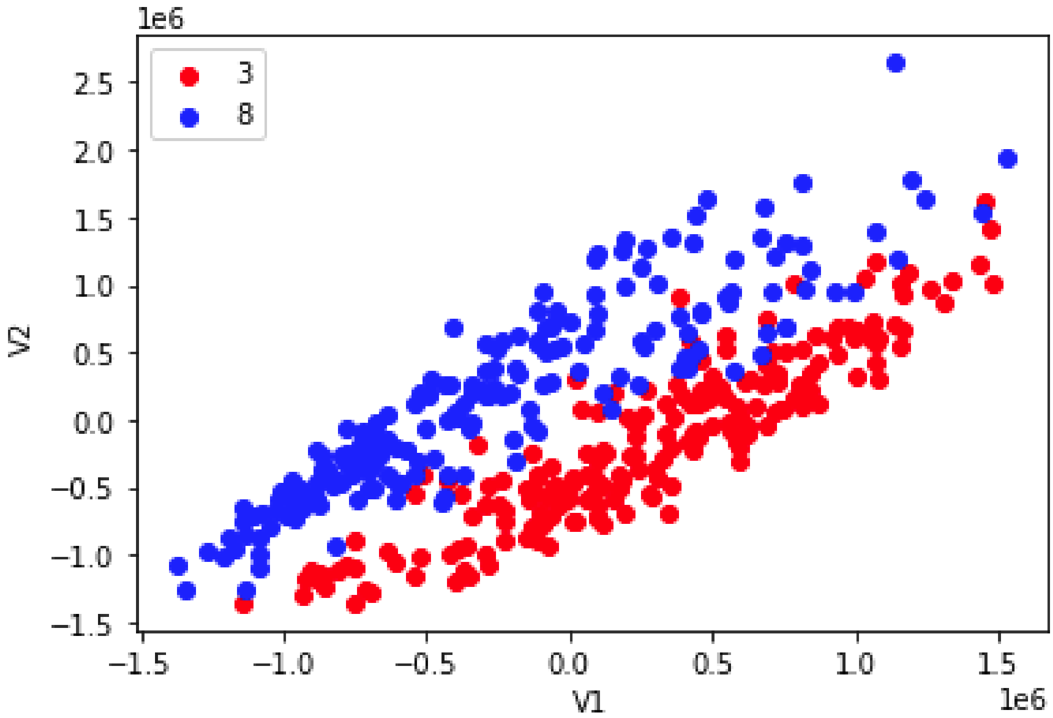
6. (10 pts) Now plot the 2D point cloud of the 400 digits after projection. For visual interest, color the points in three.txt red and the points in eight.txt blue. But keep in mind that PCA is an unsupervised learning method, and it does not know such class labels.

Sol:

2D point cloud plotted for unit vector V1 and V2



2D point cloud plotted for scaled vectors V1 and V2



2 Directed Graphical Model [20 points]

Consider the directed graphical model (aka Bayesian network) in Figure 1.

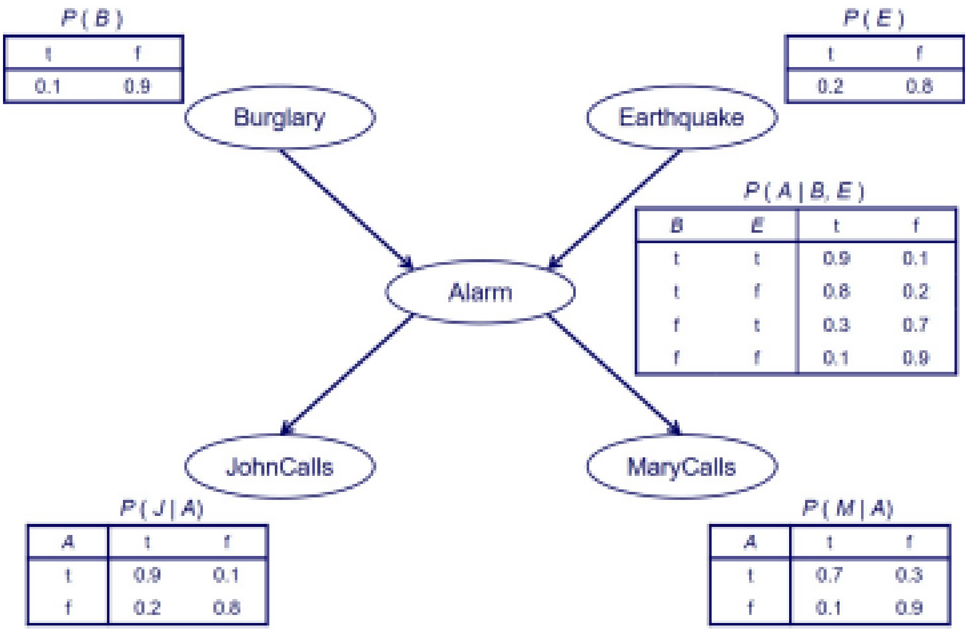


Figure 1: A Bayesian Network example.

Compute P(B = t | E = f, J = t, M = t) and P(B = t | E = t, J = t, M = t). These are the conditional probabilities of a burglar in your house (yikes!) when both of your neighbors John and Mary call you and say they hear an alarm in your house, but without or with an earthquake also going on in that area (what a busy day), respectively.

Sol:

Question 2 - Directed Graphical Model

a] $P(B=t \mid E=f, J=t, M=t) = \dfrac{P(B=t, E=f, J=t, M=t)}{P(E=f, J=t, M=t)}$

$$P(B=t, E=f, J=t, M=t) = \sum_{a, \neg a} P(B)P(E)P(A|B,E)\,P(J/A)\,P(M/A)$$

$$= 0.1 \times 0.8\,[0.8 \times 0.9 \times 0.7 + 0.2 \times 0.2 \times 0.1]$$

$$= 0.08\,[0.508] = 0.04064$$

$$P(B=f, E=f, J=t, M=t) = \sum_{a, \neg a} P(B)P(E)P(A|B,E)\,P(J/A)\,P(M/A)$$

$$= 0.9 \times 0.8\,[0.1 \times 0.9 \times 0.7 + 0.9 \times 0.2 \times 0.1]$$

$$= 0.72\,[0.081]$$

$$= 0.05833$$

$$P(E=f, J=t, M=t) = P(B=t, E=f, J=t, M=t) + P(B=f, E=f, J=t, M=t)$$

$$= 0.04064 + 0.05832$$

$$= 0.09896$$

$$\Rightarrow P(B=t \mid J=t, E=f, M=t) = \frac{0.04064}{0.04064 + 0.05832} = \boxed{0.41067}$$

b) $P(B=t \mid E=t, J=t, M=t) = \dfrac{P(B=t, E=t, J=t, M=t)}{P(E=t, J=t, M=t)}$

$$P(B=t, E=t, J=t, M=t) = \sum_{a, \neg a} P(B)P(E)P(A \mid B,E)P(J \mid A) P(M \mid A)$$

$$= 0.1 \times 0.2 \left[0.9 \times 0.9 \times 0.7 + 0.1 \times 0.2 \times 0.1\right]$$

$$= 0.02 \left[0.569\right]$$

$$= 0.01138$$

$$P(B=f, E=t, J=t, M=t) = \sum_{a, \neg a} P(B)P(E)P(A \mid B,E)P(J \mid A) P(M \mid A)$$

$$= 0.9 \times 0.2 \left[0.3 \times 0.9 \times 0.7 + 0.7 \times 0.2 \times 0.1\right]$$

$$= 0.18 \left[0.203\right]$$

$$= 0.03654$$

$$P(E=t, J=t, M=t) = P(B=t, E=t, J=t, M=t) + P(B=f, E=t, J=t, M=t)$$

$$= 0.01138 + 0.03654$$

$$= 0.04792$$

$$\Rightarrow P(B=t \mid J=t, E=t, M=t) = \frac{0.01138}{(0.01138 + 0.03654)}$$

$$= \boxed{0.237479}$$