HOMEWORK 7

Madhushree Nijagal
9084490524

Instructions: You can choose any programming language as long as you implement the algorithm from scratch. Use this latex file as a template to develop your homework. Submit your homework on time as a single pdf file to Canvas. Please check Piazza for updates about the homework.

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`.

   Visualising the first line of three.txt and eight.txt

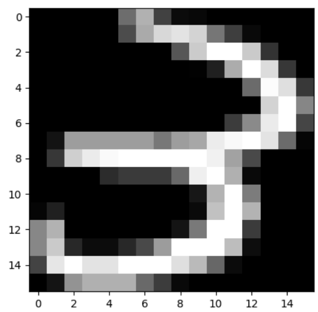


Figure 1: three.txt

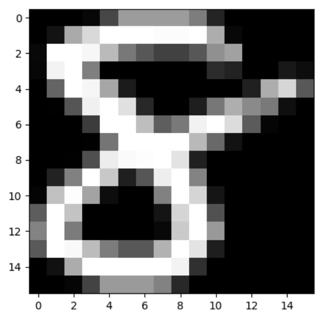


Figure 2: eight.txt

2. (10 pts) Putting the two data files together (three first, eight next) to form an $n \times D$ matrix $\mathbf{X}$ where $n = 400$ digits and $D = 256$ pixels. Note we use $n \times D$ size for $\mathbf{X}$ instead of $D \times n$ to be consistent with the convention in linear regression. The i-th row of X is $\mathbf{x}_i^\top$, where $\mathbf{x}_i \in \mathbb{R}^D$ is the i-th image in the combined data set. Compute the sample mean $\mathbf{y} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i$. Visualize $\mathbf{y}$ as a 16×16 grayscale image.

   Visualizing the mean as a 16X16 image:

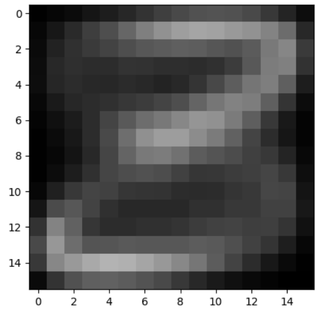


Figure 3: Sample mean of all 400 points

3. (10 pts) Center $\mathbf{X}$ using $\mathbf{y}$ above. Then form the sample covariance matrix $\mathbf{S} = \frac{\mathbf{X}^\top \mathbf{X}}{n-1}$ Show the 5×5 submatrix $\mathbf{S}(1 \ldots 5, 1 \ldots 5)$.
   The submatrix from the sample covariance matrix:

$$S[1:5,1:5] = \begin{bmatrix} 59.167 & 142.149 & 28.682 & -7.178 & -14.335 \\ 142.149 & 878.938 & 374.137 & 24.127 & -87.127 \\ 28.682 & 374.137 & 1082.905 & 555.226 & 33.724 \\ -7.178 & 24.127 & 555.226 & 1181.244 & 777.771 \\ -14.335 & -87.1278 & 33.724 & 777.771 & 1429.959 \end{bmatrix}$$

4. (10 pts) Use appropriate software to compute the two largest eigenvalues $\lambda_1 \geq \lambda_2$ and the corresponding eigenvectors $\mathbf{v}^1, \mathbf{v}^2$ of $\mathbf{S}$. For example, in Matlab one can use `eigs(S,2)`. Show the value of λ_1, λ_2. Visualize $\mathbf{v}^1, \mathbf{v}^2$ as 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.

   The first largest eigenvalue $\lambda_1 = 237155.24629048537$
   The second largest eigenvalue $\lambda_2 = 145188.352686825$

   Visualising v^1 and v^2

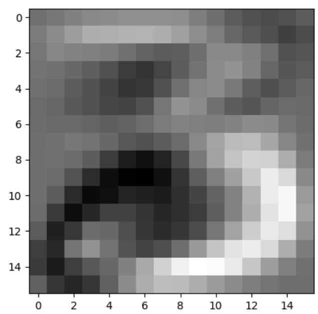


Figure 4: v1 corresponding to λ_1

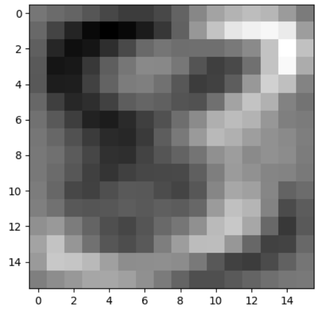


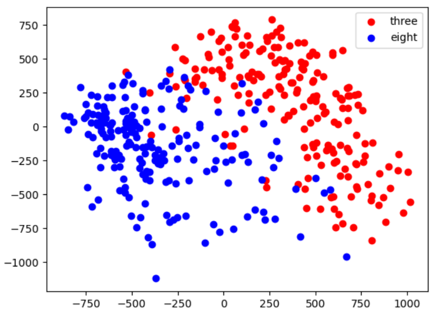
Figure 5: v2 corresponding to λ_2

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

   Coordinate from three: [136.20872784 242.62848028]
   Coordinate from eight: [-312.68702792 -649.57346086]

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.

   The 2D plot of 400 points after PCA:



2 Directed Graphical Model [20 points]

Consider the directed graphical model (aka Bayesian network) in Figure 6.

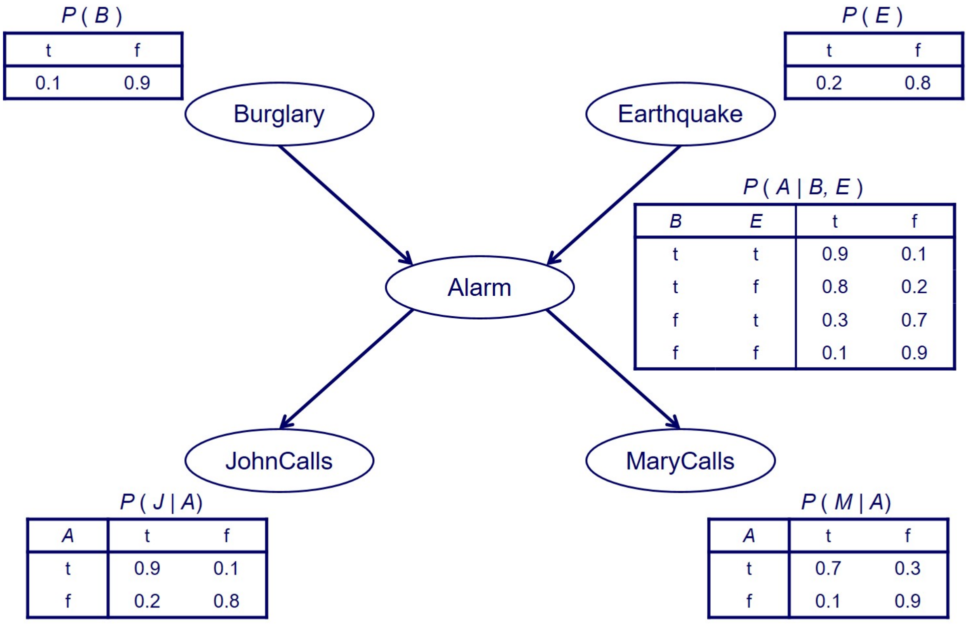


Figure 6: A Bayesian Network example.

Compute $\mathbb{P}(B = t \mid E = f, J = t, M = t)$ and $\mathbb{P}(B = t \mid E = t, J = t, M = t)$. 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.

To get the solution, the inference over enumeration needs to be done over the shown bayesian network.
We know that the conditional probability can be expressed as:

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

Calculating the numerator $P(B = t, E = f, J = t, M = t)$:

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

$= P(B = t)P(E = f)[P(A = t|B = t, E = f)P(J = t|A = t)P(M = t|A = t) + P(A = f|B = t, E = f)P(J = t|A = f)P(M = t|A = f)]$

$= (0.1 * 0.8) * [(0.8 * 0.9 * 0.7) + (0.2 * 0.2 * 0.1)]$

$= 0.04064$

Calculating the denominator $P(B = f, E = f, J = t, M = t)$:

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

$P(B = f, E = f, J = t, M = t) = P(B = f)P(E = f)[P(A = t|B = f, E = f)P(J = t|A = t)P(M = t|A = t) + P(A = f|B = f, E = f)P(J = t|A = f)P(M = t|A = f)]$

$= (0.9 * 0.8) * [(0.1 * 0.9 * 0.7) + (0.9 * 0.2 * 0.1)]$

$= 0.05832$

Finally calculating the value of $P(B = t \mid E = f, J = t, M = t)$

$P(B=t \mid E=f, J=t, M=t) = \frac{0.04064}{0.04064+0.05832} = 0.4106709$

Similarly, calculating $\mathbb{P}(B=t \mid E=t, J=t, M=t)$

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

Calculating the numerator $P(B=t, E=t, J=t, M=t)$:

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

$= P(B=t)P(E=t)[P(A=t|B=t, E=t)P(J=t|A=t)P(M=t|A=t) + P(A=f|B=t, E=t)P(J=t|A=f)P(M=t|A=f)]$

$= (0.1 * 0.2) * [(0.9 * 0.9 * 0.7) + (0.1 * 0.2 * 0.1)]$

$= 0.01138$

Calculating the denominator $P(B=f, E=t, J=t, M=t)$:

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

$P(B=f, E=t, J=t, M=t) = P(B=f)P(E=t)[P(A=t|B=f, E=t)P(J=t|A=t)P(M=t|A=t) + P(A=f|B=f, E=t)P(J=t|A=f)P(M=t|A=f)]$

$= (0.9 * 0.2) * [(0.3 * 0.9 * 0.7) + (0.7 * 0.2 * 0.1)]$

$= 0.03654$

Finally calculating the value of $P(B=t \mid E=t, J=t, M=t)$

$P(B=t \mid E=t, J=t, M=t) = \frac{0.01138}{0.01138+0.03654} = 0.2374913$