IITM-CS5691 : Pattern Recognition and Machine Learning — Release Date: October 9, 2023
Assignment II — Due Date : October 23, 2023, 23:59

Roll No: CS22Z121 — Name: Sandeep Kumar Suresh
Collaborators (if any):
References/sources (if any): PCA : Page 336 -337 Mathematics for Machine Learning
https://www.askpython.com/python/examples/principal-component-analysis
https://maths.nuigalway.ie/ rquinlan/linearalgebra/section3-1.pdf
https://math.stackexchange.com/questions/3082031/the-effect-of-adding-an-edge-on-the-laplacian-of-a-weighted-digraph
https://web.stanford.edu/class/cs168/l/l11.pdf

Solution: 1 a

$$x^\top Lx = x^\top (D - A)x = x^\top Dx - x^\top Ax$$

since D is a Diagonal Matix. Tha whole equation can be written in Summation form.

$$= \sum_{l=1}^{n} d_i {x_i}^2 - \sum_{ij=1}^{n} x_i x_j \omega ij$$

multiplying and dividing by two

$$= \frac{1}{2}\left[2\sum_{i=1}^{n} d_i x_i^2 - 2\sum_{ij=1}^{n} x_i x_j w_{ij}\right]$$
$$\text{given } d_i = d_j$$
$$= \frac{1}{2}\left[\sum_{i=1}^{n} d_i x_i^2 - 2\sum_{ij=1}^{n} x_i x_j w_{ij} + \sum_{i=1}^{n} d_j x_{jj}^2\right]$$
$$= \frac{1}{2}\sum_{ij=1}^{n} w_{ij}\left(x_i - x_j\right)^2$$

The quantity $x^\top Lx$ is the sum of squares of the differences between the values of neibouring nodes, or alternatively the sum of squares of the distance between neighbours.

- The eigenvalue of symmetric matrix with real elements are always real.
- Here D and A are symmetric matrix . Therefore the eigenvalue of L are real.
- Also here L is a symmetric matrix and $x^\top Lx \geqslant 0$ since $w_{ij} \geqslant 0$
- The weights wij are non -negative

From the above , we can say that it is positive semidefinite matrix . Therefore we can say that all the eigenvalues are non-negative and the least eigenvalue of L is zero.

b.

- If the multiplicity of eigenvalue tells us about the number of connected component in the graph . The eigenvector will have constant one vector 1 as one of the eigenvector and the other two eigenvector will be different.
- eigenvalue corresponsding to the 0 corresponds to vectors for which neighbours have similar values.
- From the above we can say that the eigenvectors with eigenvalue zero are constant on each connected component.

c.

Let $L_G = D - A$ and $L'_G = D' - A'$
Since we are adding one more node , the ajjacency matrix and degree matrix changes . Therefore the Laplacian Matrix also changes .
The multiplicity of the eigenvalue 0 of L'_G will be same or decrease to that of L_G . The added edge can connect to the previous component or can make the single component depending on where it is connected in the graph.

d.

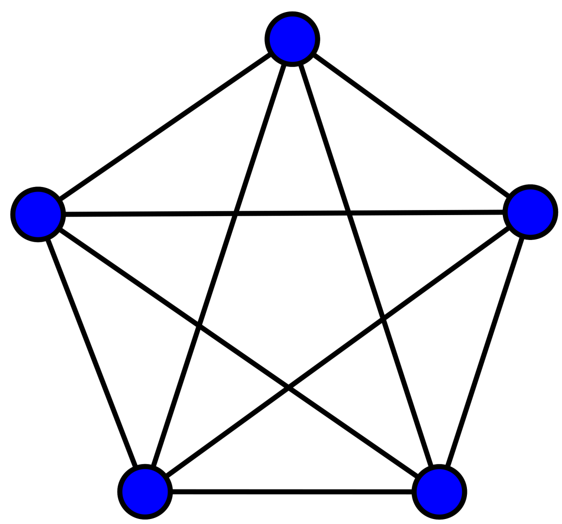

Figure 1: Complete Graph

In the above Complete Graph with 5 nodes , we can see that the there is only one connected component and the multiplicity of eigenvalue 0 will be 1 .
This can be also explained by the fact that the above graph forms only one cluster and .
Given that the above graph forms only one cluster , it is evident that the rest of the eigenvalue will be non - negative value .
$\nu = (11111)^T$ be eigenvector corresponding to eigenvalue 0 .

The Laplacian matrix L_G can be written as Eigenvector = $\begin{bmatrix} 4 & -1 & -1 & -1 & -1 \\ -1 & 4 & -1 & -1 & -1 \\ -1 & -1 & 4 & -1 & -1 \\ -1 & -1 & -1 & 4 & -1 \\ -1 & -1 & -1 & -1 & 4 \end{bmatrix}$

Let us assume $v(1)=\lambda$, where λ is the eigenvalue corresponding to eigenvector $v(1)$ Then we can say that

$L_G\ v(1) = \lambda\ v(1)$

Here $v(1)$ will be a eigenvector which will be orthogonal to v

Solution:

2 .

a) Given the PC1 $= (-0.694, -0.720)$, we noed to find the projection of data 2

$$[20.82, 24.03] \begin{bmatrix} -0.694 \\ -0.720 \end{bmatrix} = -31.75$$

The reconstructed datapoint

$$= -31.75 \begin{bmatrix} -0.694 \\ -0.720 \end{bmatrix} = \begin{bmatrix} 22.03 \\ 22.86 \end{bmatrix}$$

Reconstruction Error

$$\begin{aligned} &= (22.03 - 20.82)^2 + (22.86 - 24.03)^2 \\ &= 1.46 + 136 = 2.82 \end{aligned}$$

b)

$$PC_2 = (0.720, -0.694)$$

$$\text{Projection of data2 } = [20.82, 24.03] \begin{bmatrix} 0.72 \\ -0.694 \end{bmatrix} = -1.68$$

$$\text{Reconstruction using } PC_2 = -1.68 \begin{bmatrix} 0.72 \\ -0.694 \end{bmatrix} = \begin{bmatrix} -1.20 \\ 1.15 \end{bmatrix}$$

$$PC_1 + PC_2 = \begin{bmatrix} 20.83 \\ 24.02 \end{bmatrix}$$

Reconstruction Error

$$\begin{aligned} &= (20.83 - 20.82)^2 + (04.02 - 24.03)^2 \\ &= 0.0001 + 0.0001 \approx 0 \end{aligned}$$

c) mean $= [9.293, 9.685]$

PC_1 and PC_2 are same as above

Eigenvalue = $[0.816, 181.456]$

Eigenvector = $\begin{bmatrix} -0.719 & 0.694 \\ 0.694 & 0.719 \end{bmatrix}$

d) Eigenvalue $= [0, 0.816, 181.45]$

$$\text{Eigenvector } = \begin{bmatrix} 0 & -0.719 & 0.694 \\ 0 & 0.694 & 0.719 \\ 1 & 0 & 0 \end{bmatrix}$$

data 2 $= (00.82, 24.03, 3.5)$
Projection $= -31.75$
From the above , we can say that a constant value change the eigenvalue and eigenvector.
It will be same as from the above question.Since the z coordinate
does not play any role . The Principal Component will have two coordinates

$$\text{Reconstructed Datapoint } = \begin{bmatrix} 22.03 \\ 22.86 \end{bmatrix}$$

Reconstruction Error $= 1 \cdot 46 + 1 \cdot 36 = 2.82$

Solution:

3 a.

$$\tilde{E}(\omega) = \frac{1}{2}\left[\sum_{n=1}^{N}\left(t_n - \omega^{+}\phi\left(x_n\right)\right)^2 + \lambda\omega^{\top}\omega\right]$$

We need to minimize the inside term.

$$\begin{gathered}
\left(t_n - \phi\omega\right)^{\top}\left(t_n - \phi\omega\right) + \lambda\omega^{\top}\omega \\
\left(\omega^{\top} - \omega^{\top}\phi^{\top}\right)\left(t_n - \phi\omega\right) + \lambda\omega^{\top}\omega \\
t_n^{\top}t_n - t_n^{\top}\phi\omega - \omega^{\top}\phi^{\top}t_n + \omega^{\top}\phi\phi\omega + \lambda\omega^{\top}\omega
\end{gathered}$$

$$t_n^{\top}t_n - 2\omega^{\top}\phi_n^{\top}t_n - \omega^{\top}\phi\phi\omega + \lambda\omega^{\top}\omega \tag{1}$$

Taking the derivative of above equation 1 w.r.t to ω and assigning it to zero gives the below

$$\begin{gathered}
-2\phi^{\top}t_n - 2\Phi^{\top}\phi\omega + 2\lambda I\omega = 0 \\
\left(\phi^{\top}\phi + \lambda I\right)\omega = \phi^{\top}t_n \\
\omega = \left(\phi^{\top}\phi + \lambda I\right)^{-1}\phi^{\top}t_n
\end{gathered}$$

Taking the double derivate of equation 1, we get

$$2\lambda\omega \tag{2}$$

From equation 2 we cawhich is greater than zero . Therefore the function is convex

b.

$x = \begin{bmatrix} -2 & 4 \\ -1 & 2 \end{bmatrix} \quad t = \begin{bmatrix} 3 & -1 \end{bmatrix}$

$E(\omega) = \frac{1}{2}\|x\omega - t\|^2$

$= \frac{1}{2}(x\omega - t)^{\top}(x\omega - t)$

$$\begin{aligned}
&\omega = \left(x^{\top}x\right)^{-1}x^{\top}t \\
&x^{\top}x = \begin{bmatrix} -2 & -1 \\ 4 & 2 \end{bmatrix}\begin{bmatrix} -2 & 4 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 5 & -10 \\ -10 & 20 \end{bmatrix} \\
&\left|x^{\top}x\right| = 0
\end{aligned}$$

since determinant is zero , there will be infinite solution. We can only find a solution using the psedo-inverse

$$\widetilde{E}(\omega) = \frac{1}{2}\sum_{n=1}^{N}\left(t_n - \omega^\top \phi\left(x_n\right)\right)^2 + \frac{d}{2}\omega^\top \omega$$

$$\lambda = 1$$

$$\tilde{\epsilon}(\omega) = \frac{1}{2}\sum_{n=1}^{N}\left(t_n - \omega^\top \phi\left(x_n\right)\right)^2 + \frac{1}{2}\omega^\top \omega$$

$$\omega = \left(x^\top x + I\right)^{-1} x^\top t$$

From the above $x^\top x + I = \begin{bmatrix} 6 & -10 \\ -10 & 21 \end{bmatrix}$

$$\left|x^\top x + I\right| = 26$$

$$\left(x^\top x + I\right)^{-1} x^\top t = \tfrac{1}{26}\begin{bmatrix} 21 & 10 \\ 10 & 6 \end{bmatrix}\begin{bmatrix} -2 & -1 \\ 4 & 2 \end{bmatrix}\begin{bmatrix} 3 \\ -1 \end{bmatrix} = \tfrac{1}{26}\begin{bmatrix} -5 \\ 10 \end{bmatrix}$$

Using Regularised Least Square minimiser , we would get a unique solution instead of an infinite solution in the first case.

Solution: 4 .
For the below Question i have plotted some of the extra graph for more intuitive understanding of what the data looks like.
(a)

Figure 2: data without mean centering

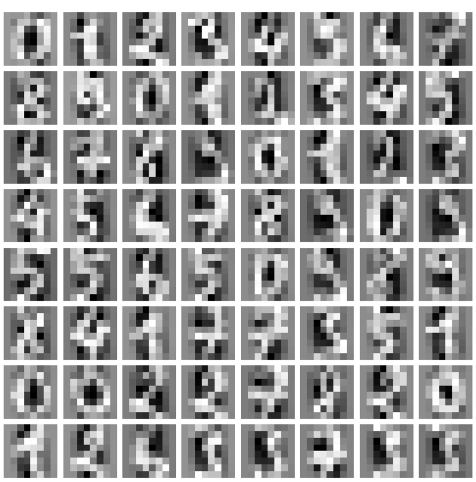

Figure 3: Data with mean centering

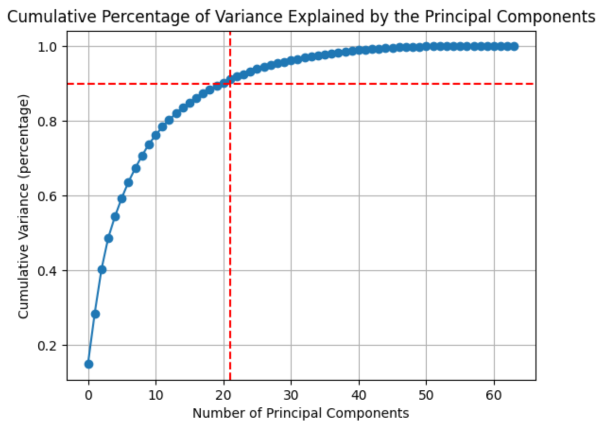

Figure 4: Number of principal components that contribute to 90 of the variance in the dataset.

(b)

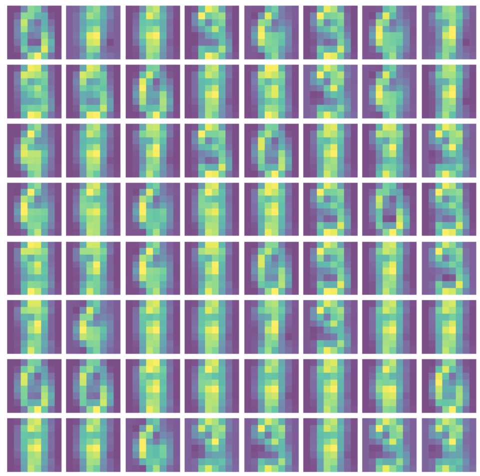

Figure 5: Reconstruction of the data with the 2 PCA Components.png

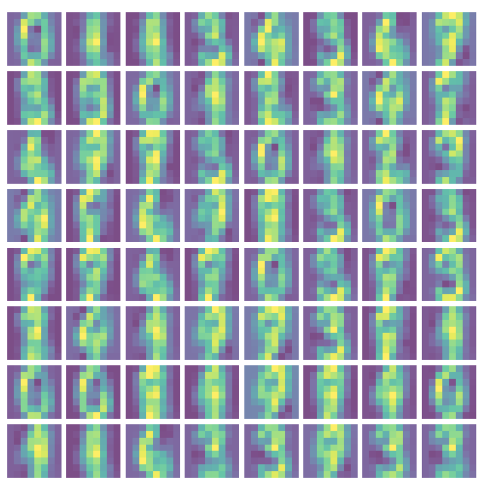

Figure 6: Reconstruction of the data with the 4 PCA Components.png

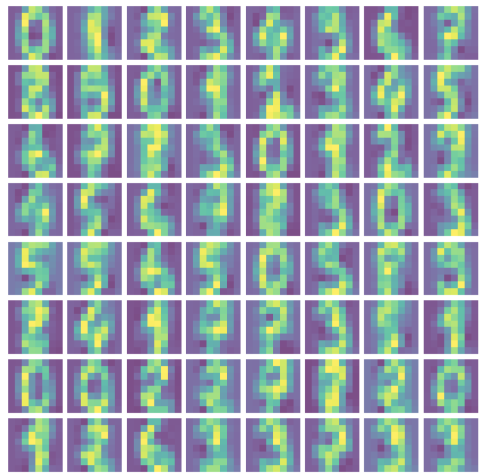

Figure 7: Reconstruction of the data with the 8 PCA Components.png

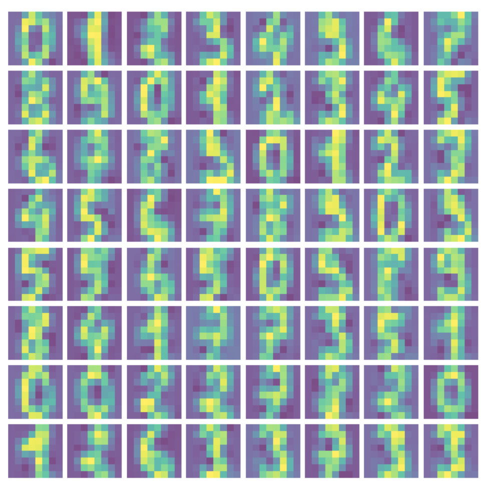

Figure 8: Reconstruction of the data with the 16 PCA Components.png

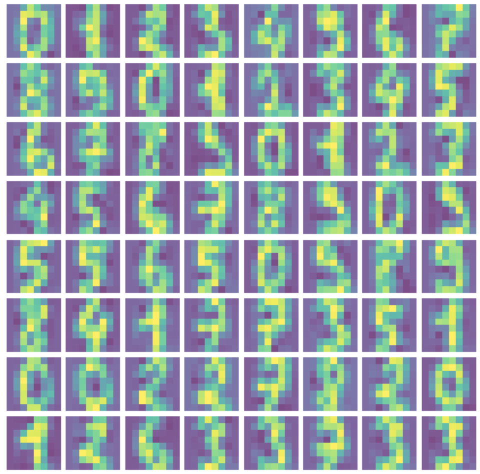

Figure 9: Reconstruction of the data with the 21 PCA Components.png

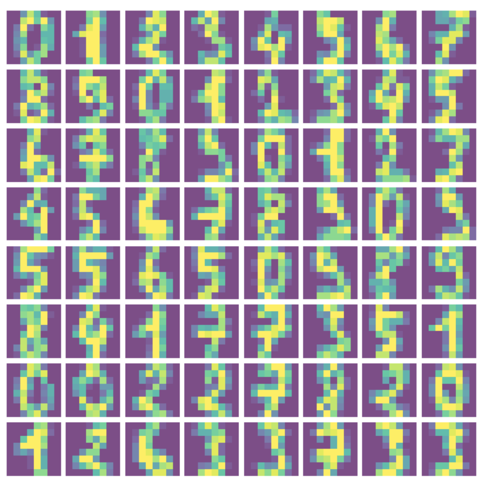

Figure 10: Reconstruction of the data with the 64 PCA Components.png

Dimension	**Mean-Squared-Error**
2	13.4210
4	9.6280
8	6.1218
16	2.8272
21	1.8173
64	0.0000

Table 1: MSE wrt. the number of PCA Component

Here i have taken the raw data given without mean centering and the reconstructed data after adding the mean to the data.

For the Question asked in the assignment , the optimal dimension would be 16. For my better understanding did the MSE for 21 and 64 Components.

Here we can observe that the as the number of PCA component increases , the mean squared error decreases and we can observe that the error is relatively small for if we reconstruct the data with 16 components.Taking all the 64 eigenvectors we get a zero MSE which is theoretically correct

We can also observe that the Reconstruction with 21 components gives a MSE close to zero. Ideally the dimensions should be selected based on maximising variance and minizing MSE.