Homework Assignment 10 [30 points]

Simon Lizarazo
simonl3

STAT437 Unsupervised Learning – Fall 2023

Due: Friday, November 3 on Canvas at 11:59pm CST.

- **Answer questions 1-4 in** assignment_10.ipynb
- **Answer questions 5-7** in this pdf below

Problem	Points
1.1.1	0.5
1.1.2	0.75
1.2	0.75
2.1	0.75
2.2	0.75
3.1	1.5
3.2	1
4.1	1
4.2.1	1
4.2.2	1
4.2.3	1.5
4.3.1.	1
4.3.2	0.75
4.3.3	0.75
4.3.4	0.75
4.3.5	1
4.3.6	1
4.3.7	1.5
4.4.1	0.75
4.4.2	0.75
4.4.3	0.75
4.4.4	1
4.4.5	1
4.4.6	1
5	3.5
6	1
7	3

Question 5

A partially completed, mean-centered, $X_{5\times4}$ matrix is given below. It's covariance matrix $C_{4\times4}$ is also given below. Finally, the four eigenvectors of $C_{4\times4}$ and their corresponding eigenvalues are given below. Note that these eigenvalues are not in order.

We decide to use PCA to project $X_{5\times4}$ onto the matrix $Y_{5\times2}$ (ie. two principal components) also shown below and partially completed. Use the information given to fill in the blanks for the 7 green boxes shown below. Explain your answers or show your work.

The sum of Eigenvalues is = to the ΣVar : = 4.19

C: Covariance Matrix of X				
	Attribute 1	Attribute 2	Attribute 3	Attribute 4
Attribute 1	1.2	0.8	0.4	0
Attribute 2	0.8	1.19	0.6	0.5
Attribute 3	0.4	0.6	0.8	0.75
Attribute 4	0	0.5	0.75	1

$\Sigma Var = 4.19 = 1.2 + x + 0.8 + 1$

$x = 4.19 - 3$

$x = 1.19$

Eigenvalues of C			
Eigenvalue 1	Eigenvalue 2	Eigenvalue 3	Eigenvalue 4
0.30	1.22	2.62	0.05
3	2	1	4

Eigenvectors of C			
Eigenvector 1	Eigenvector 2	Eigenvector 3	Eigenvector 4
0.47	0.66	0.48	0.35
-0.75	0.17	0.61	-0.15
0.45	-0.31	0.48	-0.68
0.05	-0.67	0.41	0.62

Original Mean-Scaled Dataset X				
	Attribute 1	Attribute 2	Attribute 3	Attribute 4
Object 1	-1.2	-0.8	0.6	1
Object 2	?	?	?	?
Object 3	?	?	?	?
Object 4	?	?	?	?
Object 5	?	?	?	?

Principal Components Y		
	Principal Component 1	Principal Component 2
Object 1	-0.366	-1.784
Object 2	?	?
Object 3	?	?
Object 4	?	?
Object 5	?	?

Object 1 PC_1

= object 1 PC x Eigenvector 3

$= -1.2 \times 0.48 - 0.8 \times 0.61 + 0.6 \times 0.48 + 1 \times 0.41$

$= -0.366$

Object 1 PC_2

= object 1 PC x Eigenvector 2

$= -1.2 \times 0.66 - 0.8 \times 0.17 - 0.6 \times 0.31 - 0.67 \times 1$

$= -1.784$

Covariance Matrix of Y		
	Principal Component 1	Principal Component 2
Principal Component 1	2.62	0
Principal Component 2	0	1.22

The variance is equal to the highest eigenvalue for PC1 and the second highest for PC2.

orthogonal to each other then no covariance.

Question 6

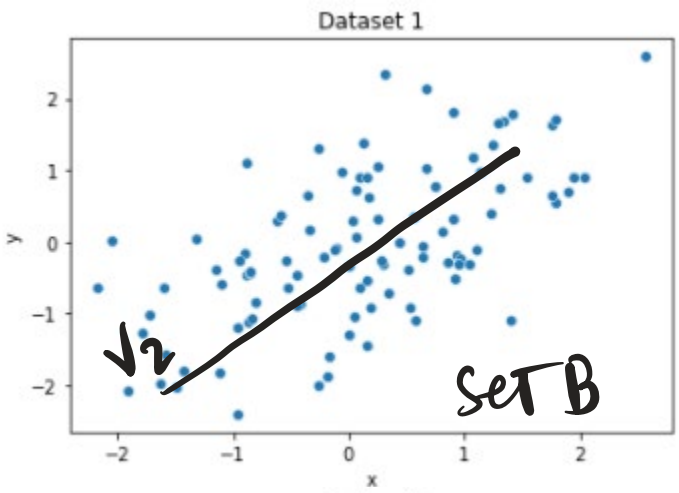
What percent of total original attribute variability is preserved in the two principal components above?

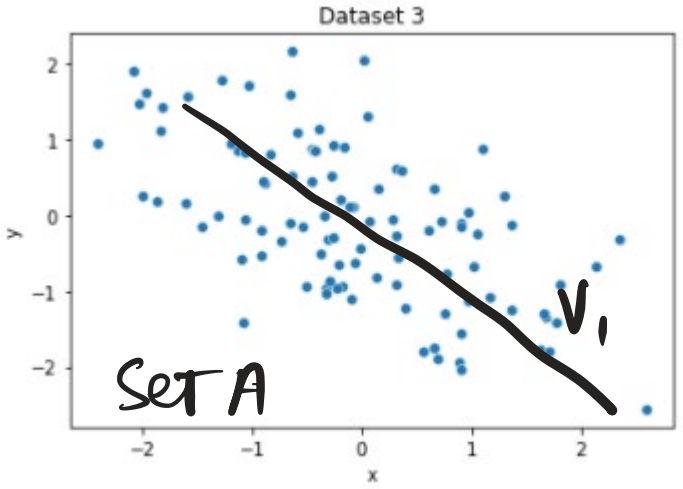
$\frac{2.62}{4.19} = 0.62\ (PC_1)$ $\frac{1.22}{4.19} = 0.30\ (PC_2)$. $PC_1 + PC_2 = 0.92$.

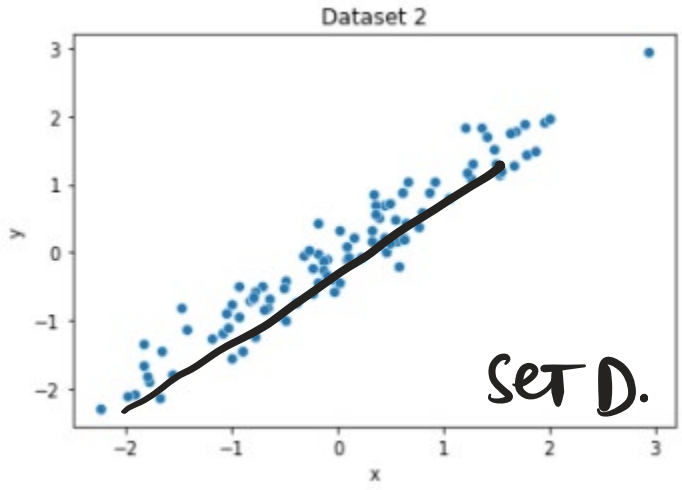
92% of the original variance

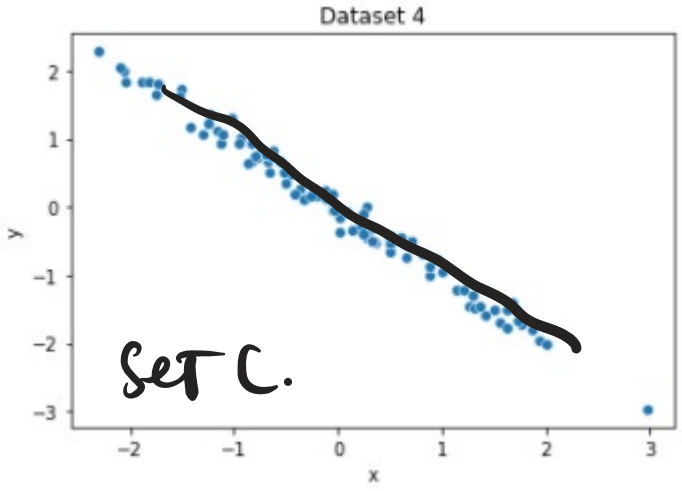
Question 7

Each of the datasets below have a mean of [0,0]. The covariance matrix was calculated for each of these datasets. Then, the corresponding eigenvalues and eigenvectors for each of these 4 covariances matrices were calculated and are shown below in A-D. Match each dataset (1-4) to its corresponding set of eigenvectors and eigenvalues (A-D).









Set A

eigenvalue_1	eigenvalue_2
1.808116	0.439048

eigenvector_1	eigenvector_2
0.726797	0.686853
-0.686853	0.726797

Set B

eigenvalue_1	eigenvalue_2
0.439048	1.808116

eigenvector_1	eigenvector_2
-0.726797	-0.686853
0.686853	-0.726797

Set C

eigenvalue_1	eigenvalue_2
2.397335	0.008786

eigenvector_1	eigenvector_2
0.708968	0.705240
-0.705240	0.708968

Set D

eigenvalue_1	eigenvalue_2
0.04393	2.349215

eigenvector_1	eigenvector_2
-0.711369	-0.702819
0.702819	-0.711369