I TD on CPA

1. Quiz: TD 2

exercise 1 : Introduction to Centering, Scaling, and PCA Analysis

The following data represents the dollar sales of different product categories A, B, C across three regions R_1, R_2, R_3:

Production (PR)	Region R_1	Region R_2	Region R_3
A	200	220	240
β	150	180	190
C	300	310	320

Let X represent the data matrix corresponding to the above table:

$$X = \begin{bmatrix} 200 & 220 & 240 \\ 150 & 180 & 190 \\ 300 & 310 & 320 \end{bmatrix}.$$

Answer the following questions to explore the concepts of centering, scaling, and preparing the data for Principal Component Analysis (PCA):

Question 1

1. Calculate the ***gravity center*** g of the data matrix X (g is the vector of column means of X $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_{i\cdot}$,.
2. Compute the vector σ, where each element σ_i is the standard deviation of the column X_i: $\sigma_i = \sqrt{\frac{1}{n} \sum_{k=1}^{n} (X_{ki} - g_i)^2}$.
3. Construct the weighted matrix $D(1/\sigma)$, where:
$$D(1/\sigma) = \begin{bmatrix} 1/\sigma_1 & 0 & \cdots & 0 \\ 0 & 1/\sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/\sigma_p \end{bmatrix}.$$
4. Calculate Z, the ***centered version*** of X: $Z = X - \mathbf{1}_n g^\top$, where $\mathbf{1}_n$ is an $n \times 1$ vector of ones. What is the mean of each column $Z_{\cdot j}$ of the matrix Z?
5. Compute Z^*, the ***centered and reduced version*** of X:
$Z^* = ZD(1/\sigma)$,
or equivalently:
$Z^* = (X - \mathbf{1}_n g^\top)D(1/\sigma)$.
6. Deduce the variance of each column $Z_{\cdot j}$ of Z^*.
7. Calculate Σ, the covariance matrix of Z: $\Sigma = \frac{1}{n} Z^\top Z$.
8. Recall the steps of Principal Component Analysis (PCA) based on this situation.

Exercise 2: Perform a Principal Component Analysis (PCA)

This exercise is homework to be completed and submitted between November 25 and 29, and will be graded as part of your evaluation ! ! !

On the following matrix, starting from its dispersion matrix (data are centered but not scaled):

$$\begin{pmatrix} 2 & 2 \\ 6 & 2 \\ 6 & 4 \\ 10 & 4 \end{pmatrix}$$

Question 2

Perform a Principal Component Analysis (PCA)

Hint:

> ***### Tasks:***
>
> *1. Compute the covariance matrix of the data.*
>
> *2. Calculate the eigenvalues and eigenvectors of the covariance matrix.*
>
> *3. Determine the principal components and interpret their directions.*
>
> *4. Project the original data onto the new principal component axes.*