

1. Consider a data vector

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}$$

Recall that the variance of  $X$  is defined by the formula

$$\text{Var}(X) = \frac{1}{N} \sum_{i=1}^N (x_i - m_X)^2$$

where  $m_X = \frac{1}{N} \sum_{i=1}^N x_i$  is the mean of  $X$ . Show that

$$\text{Var}(X) = \frac{1}{2N^2} \sum_{1 \leq i, j \leq N} (x_i - x_j)^2$$

2. Consider the following demeaned data matrix:

$$A = \begin{bmatrix} -27 & 38 & 37 \\ 8 & 22 & 0 \\ 14 & -25 & -35 \\ -19 & -23 & -16 \\ 36 & 32 & 14 \\ 34 & 39 & 33 \\ -40 & -24 & -24 \\ -14 & -3 & -34 \\ -2 & -30 & 3 \\ 10 & -26 & 22 \end{bmatrix}$$

a) Compute the first, second and third principal axes of  $A$ .

b) Compute the first principal component of  $A$ .

c) What is the total variance of  $A$ ?

d) What fraction of the total variance of  $A$  is captured by the first principal component? What fraction of the total variance is captured by the first two principal components?

**Note.** Use Python for computations in this problem. See the Jupyter notebook provided with this assignment for a demonstration of some tools that may be useful.

You don't need to include screenshots of the Python script, but you should explain how you performed these computations, i.e. what you computed to get each answer.

Enter your answers with precision up to two decimal digits.