Practical Session 1: Dimensionality Reduction Course

Principal Components Analysis

1 Method

Step 1 Mean normalization.

Step 2 Compute Covariance matrix.

Step 3 Compute eigenvectors/values.

Step 4 Select top $k-$ eigenvalues and their eigenvectors.

Step 5 Create an orthogonal base with the eigenvectors.

Step 6 Transform data by multiplying with said base.

2 Example

Applying the PCA method to find the transformed data for the following dataset:
$D = \{(126, 78), (128, 80), (128, 82), (130, 82), (130, 84), (132, 86)\}$.

Step 1 Mean normalization.
$\bar{x} = \cdots, \bar{y} = \cdots$
$\bar{A} =$

Step 2 Compute Covariance matrix.
$Cov(\bar{A}) = \cdots$

Step 3 Compute eigenvectors/values.
$det(\bar{A} - \lambda\mathbb{I}) = 0$
$\lambda = \cdots$

Step 4 Select top $k-$ eigenvalues and their eigenvectors.
$eig_1 = (x_1, y_1), eig_2 = (x_2, y_2)$
eigenvectors must be unit vectors

Step 5 Create an orthogonal base with the eigenvectors.
$V = (eig_1, eig_2)$

Step 6 Transform data by multiplying with said base.
Transformed data $= \bar{A} * V$

3 Exercise

3.1 Maximum variance formulation

Consider a data set of observations $\{x_n\}$ where $n = 1, ..., N$ and x_n is a Euclidean variable with dimensionality D. Our goal is to project the data onto a space having dimensionality $M \leqslant D$ while maximizing the variance of the projected data.
Give detailed proof for PCA using maximum variance formulation with $M = 2$. Address the following questions:

a) State the optimization problem to maximum variance, indicating the form of the Covariance matrix.

b) Using the Lagrange function to solve the problem - finding the formula of u_1, u_2 (two principal components).

3.2 Write your calculations

Applying the PCA method to find the transformed data for the following dataset. Check and compare the result to Python code and scikit-learn libraries, visualize the transformed data and give your comments.
$D = \{(1, 0), (2, 0), (3, 0), (5, 6), (6, 6), (7, 6)\}$