In this assignment, you will learn how to perform linear discriminative analysis for dimensionality reduction. Display numbers in two-digit precision. Set the random number to 5805.

Phase I (Non python)

1. Consider two classes, ω1(5×2) and ω2(6×2). The samples are:
    a. Sketch the scatter plot using two different symbols on the x-y cartesian coordinate.
    b. Calculate the between-class variance and within-class variance. Show all your work to receive credit. (No need to standardize)

$$\omega_1 = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 3 \\ 4 & 5 \\ 5 & 5 \end{bmatrix}, \qquad \omega_2 = \begin{bmatrix} 4 & 2 \\ 5 & 0 \\ 5 & 2 \\ 3 & 2 \\ 5 & 3 \\ 6 & 3 \end{bmatrix}$$

2. Using the spectral decomposition of the Fisher criterion, find the direction that maximizes the separation between the two classes.
    a. Plot the eigenvectors on the x–y Cartesian plane, and report which eigenvector correspond to which eigenvalue obtained from the spectral decomposition.
    b. Display the data projected onto these eigenvectors.

3. Reduce the dimensionality by projecting the original data onto the lower-dimensional subspace.
    a. Plot the probability density functions of the projected data for both classes. Which direction do you choose for the projection, and why? Explain your answer.

4. Calculate the L_2 (Euclidean) distance between the two class mean before applying LDA and compare it with the L_2 distance between the two class means after projection. Show all your work.

Phase II (python)

5. Load the politician face dataset from the **sklearn** package with min_faces_per_person = 17. This is a labeled dataset where the target values are represented by numbers, and each number corresponds to a politician's image. Answer the following questions:
    a. How many unique politicians (classes) are in the dataset?
    b. How many total observations (images) are in the dataset?

    c. What is the dimensionality of the dataset?

    d. List the names of unique politician included in the dataset.

```
from sklearn.datasets import fetch_lfw_people
```

6. Display the first twenty politician images in a **5×4 subplot**.
   a. Set the figure size to (8, 10).

7. Suppose you want to perform image classification, where each politician represents a separate class.
   a. How many classes would this problem require?
   b. Is the dataset balanced? Create a plot to justify your answer.

8. Standardize the data matrix and perform **Linear Discriminant Analysis (LDA)** to reduce the dimensionality.
   a. Compute the **within-class scatter matrix (SW)** and display the first 5×5 block on the console.
   b. Compute the **between-class scatter matrix (SB)** and display the first 5×5 block on the console.

   (Hint: Use the formulas from the lecture notes for SW and SB. You are not allowed to use built-in functions except .mean().)

9. Using the spectral decomposition of the Fisher criterion, compute the eigenvalues and eigenvectors of $S_W^{-1} \, S_B$.
   a. Display the **five largest eigenvalues** and their corresponding eigenvectors.
   b. Display the **five smallest eigenvalues** and their corresponding eigenvectors.