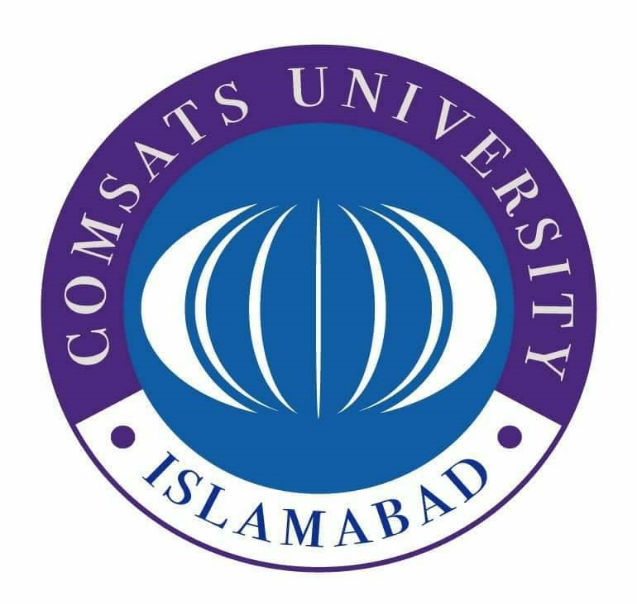
**DEPARTMENT OF COMPUTER SCIENCE, COMSATS UNIVERSITY ISLAMABAD**



**Machine Learning**

**Class Assignment 03**

**Khadija Jumani**

**(SP23-BDS-023)**

**Submission Date:**

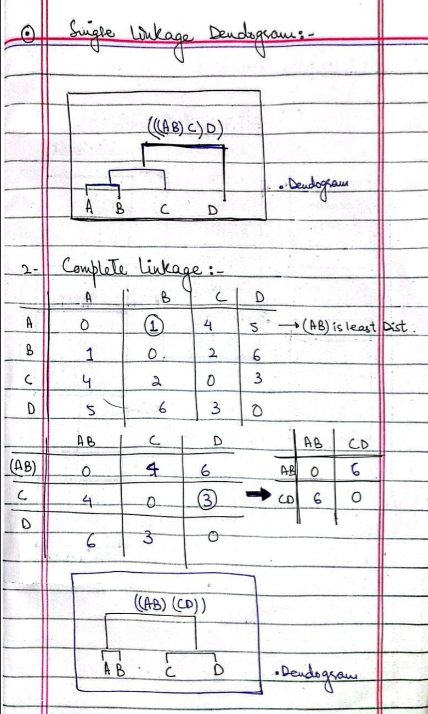
**12/10/2025**

**A table with numbers and letters

AI-generated content may be incorrect.**

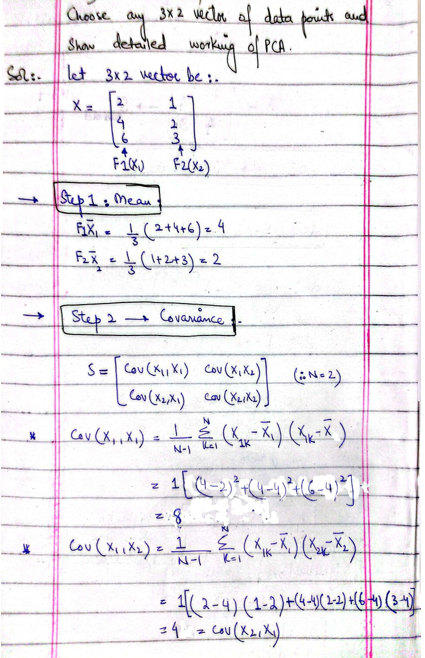
**A close-up of a notebook

AI-generated content may be incorrect.**

****

**Question.2: [10=5+5]**

1. **Choose any 3X2 vector of data points and show detailed working of PCA.**



A paper with writing on it

AI-generated content may be incorrect.

A paper with writing on it

AI-generated content may be incorrect.

A notebook with writing on it

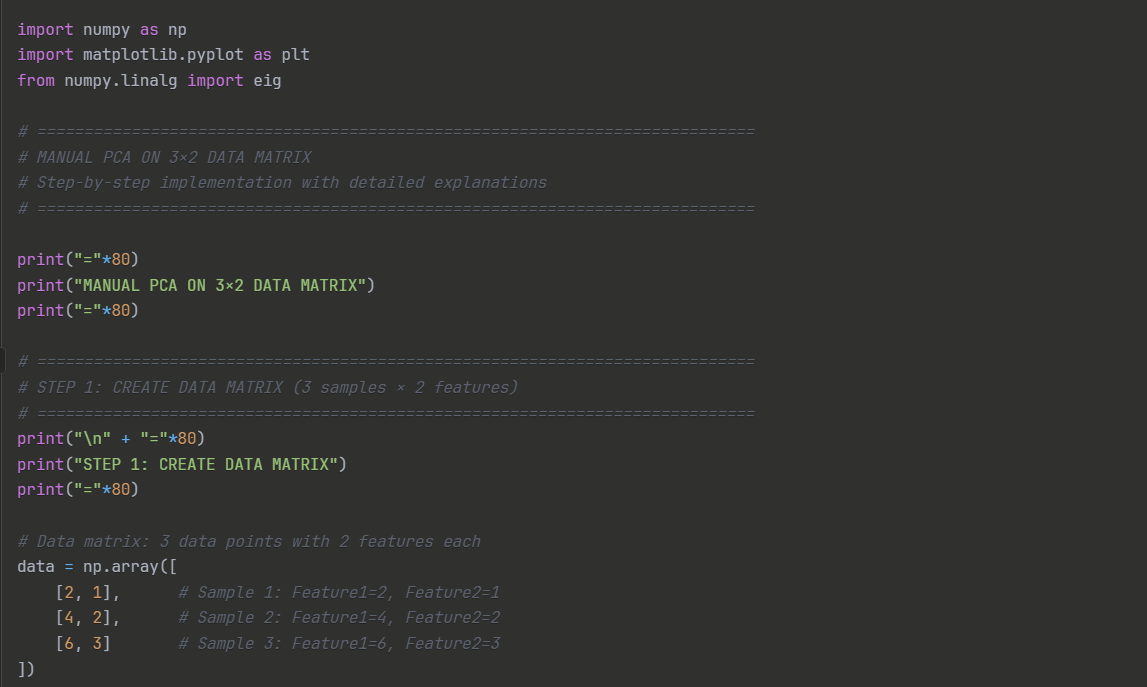
AI-generated content may be incorrect.

A graph on a notebook paper

AI-generated content may be incorrect.

1. **What This Code Does:**

**Step 1:** Creates 3×2 data matrix  
 **Step 2:** Mean centering (standardization)  
 **Step 3:** Calculates covariance matrix  
 **Step 4:** Computes eigenvalues & eigenvectors  
 **Step 5:** Calculates variance explained  
 **Step 6:** Projects data onto principal components  
 **Step 7:** Reduces from 2D to 1D  
 **Step 8:** Reconstructs original data  
 **Step 9:** Creates 4 visualizations



A screenshot of a computer program

AI-generated content may be incorrect.

A screenshot of a computer program

AI-generated content may be incorrect.

A screenshot of a computer program

AI-generated content may be incorrect.

A screenshot of a computer program

AI-generated content may be incorrect.

A screen shot of a computer program

AI-generated content may be incorrect.

A screenshot of a computer program

AI-generated content may be incorrect.

A screenshot of a computer program

AI-generated content may be incorrect.

A screen shot of a computer program

AI-generated content may be incorrect.

A computer screen shot of a program

AI-generated content may be incorrect. A computer screen shot of a program code

AI-generated content may be incorrect.

A screen shot of a computer

AI-generated content may be incorrect.

A screenshot of a computer program

AI-generated content may be incorrect.

**Output:**

**A screenshot of a computer

AI-generated content may be incorrect.**

**A screen shot of a computer

AI-generated content may be incorrect.**

**A graph of different types of data

AI-generated content may be incorrect.**

**A graph of a war zone

AI-generated content may be incorrect.**

**A screenshot of a computer

AI-generated content may be incorrect.**

1. **Get Iris.cv data set from kaggle and execute PCA on it to reduce its dimensions.**

**Overview**

The Iris dataset is one of the most famous machine learning datasets, containing 150 samples with 4 features.

**Dataset Characteristics:**

* **Number of Samples:** 150
* **Number of Features:** 4
  1. Sepal Length (cm)
  2. Sepal Width (cm)
  3. Petal Length (cm)
  4. Petal Width (cm)
* **Number of Classes:** 3 (Setosa, Versicolor, Virginica)

**Python Code for Iris PCA**

A screenshot of a computer program

AI-generated content may be incorrect.

A screen shot of a computer program

AI-generated content may be incorrect.

A screen shot of a computer program

AI-generated content may be incorrect.

A screen shot of a computer program

AI-generated content may be incorrect.

A screen shot of a computer program

AI-generated content may be incorrect.

A screen shot of a computer program

AI-generated content may be incorrect.

**Output:A computer screen with white text

AI-generated content may be incorrect.**

**A screen shot of a computer

AI-generated content may be incorrect.**

**A graph of different types of data

AI-generated content may be incorrect.**

**IMAGE 1: SCREE PLOT (Left)**

**Message:** PC1 captures 73% variance, PC2 captures 23%. After PC2, remaining components are useless (<5%). **Decision:** Use only 2 components (they cover 96% of all information).

**IMAGE 1: CUMULATIVE VARIANCE (Right)**

**Message:** Adding PC2 crosses the 95% threshold. You get 95.81% of information with just 2 out of 4 dimensions. **Decision:** Stop at PC2; no need for PC3 and PC4.

****

**IMAGE 2: 2D PCA PLOT**

**Message:**

* Setosa (red) = completely separated on LEFT (small petals)
* Versicolor (green) = middle region (medium petals)
* Virginica (blue) = right side (large petals)
* X-axis (PC1) separates by **petal size**; Y-axis (PC2) shows **sepal width variation**

**A graph of a bar graph

AI-generated content may be incorrect.**

**IMAGE 3: FEATURE LOADINGS**

**PC1 (Left Chart):** Petal width (0.95) and petal length (0.85) dominate. PC1 = "Petal size index."

**PC2 (Right Chart):** Sepal width (0.93) dominates. PC2 = "Sepal width index."

A screenshot of a computer

AI-generated content may be incorrect.

**Key Insights from Iris PCA**

1. **PC1 (72.96% variance):** Primarily represents petal length and petal width. This PC separates Setosa (small petals) from Versicolor and Virginica (larger petals).
2. **PC2 (22.85% variance):** Primarily represents sepal features. This PC helps distinguish between Versicolor and Virginica.
3. **Dimensionality Reduction:** Using only 2 principal components retains 95.81% of variance while reducing from 4 dimensions.
4. **Separation:** The 2D visualization shows clear separation between Setosa and the other two classes. Versicolor and Virginica overlap slightly but are still distinguishable.
5. **Practical Benefit:** Reduces computational complexity by 50% while maintaining most information for classification tasks.