1)

Principal Component Analysis (PCA) is a widely used dimensionality reduction technique that transforms data from a high-dimensional space to a lower-dimensional space while retaining the most important features of the data. Here's a brief explanation of the principle and the steps involved:

Principle:

The goal of PCA is to find the principal components of a dataset, which are directions of maximum variance in the data. These principal components form a new coordinate system. The idea is to project the data along these components in such a way that most of the original variance is retained in the lower-dimensional space.

Steps:

1. **Standardize the Data**: First, standardize the data so that each feature has a mean of 0 and a variance of 1. This step ensures that features with different units or scales do not disproportionately affect the result.
2. **Compute the Covariance Matrix**: Calculate the covariance matrix of the standardized data, which shows the relationships (covariances) between different features. The covariance matrix is symmetric and has dimensions equal to the number of original features.
3. **Compute Eigenvalues and Eigenvectors**: Perform eigenvalue decomposition of the covariance matrix to obtain its eigenvalues and eigenvectors. Eigenvalues represent the amount of variance explained by each principal component, and eigenvectors represent the direction of each principal component.
4. **Sort Eigenvalues and Eigenvectors**: Sort the eigenvalues in descending order and reorder the eigenvectors accordingly. The eigenvectors associated with larger eigenvalues represent the principal components that capture the most variance in the data.
5. **Select Principal Components**: Choose the top k principal components based on the largest eigenvalues. Typically, k is smaller than the original number of dimensions, and these components represent the most important features of the data.
6. **Project Data onto New Principal Components**: Finally, project the original data onto the selected k principal components. This is done by taking a linear combination of the original features based on the selected eigenvectors, resulting in the reduced-dimensional data.

By following these steps, PCA reduces the data's dimensionality while retaining as much of the original variability as possible.

2)

Problem Analysis:

Given that the variance explained by the first two principal components is 72% and 23% respectively, we need to determine the number of dimensions to retain after dimensionality reduction based on this information.

Steps:

1. Calculate the Cumulative Explained Variance Ratio:

   The variance explained by Principal Component 1 is 72%.

   The variance explained by Principal Component 2 is 23%.

   By summing the variance explained by the first two principal components, we get:

$$72\% + 23\% = 95\%$$

2. Select the Appropriate Number of Dimensions:

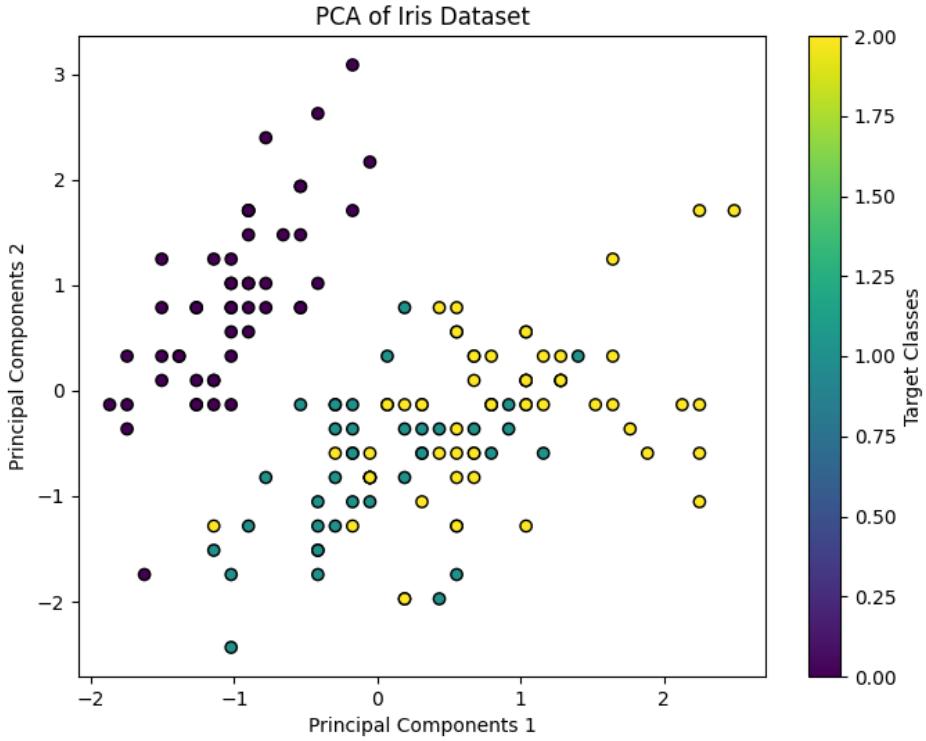
The goal of PCA is typically to retain as much variance information as possible in the data. Generally, the number of dimensions to retain is chosen such that the cumulative explained variance ratio is sufficiently high, often set at 90% or higher. This ensures that the reduced-dimensional data still represents most of the information from the original data.

3. Judgment:

The first two principal components already explain 95% of the variance, which means the remaining principal components (e.g., the third and fourth) explain only the remaining 5% of the variance.

Therefore, if we aim to retain most of the variance information, it is appropriate to keep the first two principal components (i.e., reduce the data to 2 dimensions).

The specific dimensionality reduction process can be viewed in section 2.py, and the image after dimensionality reduction is shown below.



3)

Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are both dimensionality reduction techniques, but they differ significantly in their goals and mathematical assumptions.

1. Dimensionality Reduction Goals:

- **PCA**:
    - **Goal**: PCA aims to reduce the dimensionality of the data by preserving the variance (or information) in the data. It identifies the directions (principal components) that maximize the variance in the data, regardless of class labels.
    - **Focus**: It does not take into account any class labels, focusing purely on the structure of the data (unsupervised technique).
    - **Usage**: PCA is used for tasks like feature extraction, noise reduction, and general exploratory data analysis, where the objective is to retain as much information as possible.
- **LDA**:

- **Goal**: LDA seeks to reduce the dimensionality while maximizing the separability between different classes. It identifies the directions that best separate the classes by maximizing the ratio of between-class variance to within-class variance.
- **Focus**: LDA is a supervised technique that utilizes class labels to find the best directions for class separation.
- **Usage**: LDA is often used in classification tasks where the goal is to project the data into a lower-dimensional space while preserving class separability for improved classification performance.

2. Mathematical Assumptions:

- **PCA**:
  - **Assumptions**: PCA makes fewer assumptions about the data:
    - It assumes that the directions of maximum variance (principal components) are the most important.
    - It doesn't assume any distribution of the data or class structure.
    - PCA operates under the assumption that the most informative features are those with the largest variance.
  - **Method**: It computes the eigenvectors and eigenvalues of the covariance matrix of the data, and the principal components correspond to the eigenvectors with the largest eigenvalues.
- **LDA**:
  - **Assumptions**: LDA makes more specific assumptions about the data:
    - The data for each class is assumed to follow a **Gaussian (normal) distribution**.
    - Each class is assumed to have the **same covariance matrix** (homoscedasticity).
    - The features are assumed to be **linearly separable** to some extent.
  - **Method**: LDA maximizes the between-class scatter matrix relative to the within-class scatter matrix to find the linear combinations of features that best separate the classes.

Summary of Key Differences:

Aspect	PCA	LDA
Objective	Maximize variance (unsupervised)	Maximize class separability (supervised)
Type	Unsupervised	Supervised
Data Assumptions	No assumption about class labels	Assumes Gaussian distribution for each class and same covariance matrix for all classes
Focus	Preserving overall data variance	Maximizing class separability
Methodology	Eigenvalue decomposition of covariance matrix	Maximizing ratio of between-class variance to within-class variance

In short, **PCA** is focused on capturing the most information (variance) from the data without considering any class labels, while **LDA** is focused on finding directions that maximize class separability, making it more suitable for classification tasks.