4.2. Experiment No. 1

Aim:

To use LDA (Linear Discriminant Analysis) for dimensionality reduction. You have a dataset that includes measurements for different variables on wine (alcohol, ash, magnesium, etc.). Apply the LDA algorithm and transform this data so that most variations relevant for distinguishing between red and white wines are captured by a small number of discriminant components.

Dataset Link: https://media.geeksforgeeks.org/wp-content/uploads/Wine.csv

Objective:

1. Understand and implement the LDA algorithm for dimensionality reduction and classification.
2. Transform the wine dataset to a lower-dimensional space using LDA.
3. Visualize the transformed data to observe the separation between red and white wines based on discriminant features.

Theory:

Linear Discriminant Analysis (LDA) is a supervised dimensionality reduction technique that finds the linear combinations of features which best separate two or more classes. Unlike PCA, which focuses on capturing variance, LDA focuses on maximizing class separability.

Goals of LDA:

- Reduce dimensionality while preserving as much class-discriminatory information as possible.
- Find new axes (linear discriminants) that maximize between-class variance and minimize within-class variance.
- Enhance the ability to distinguish between multiple classes.

When to Use LDA:

- When data has class labels (supervised).
- When the goal is classification or class separability rather than general variance.

- When classes overlap in the original feature space but may be separated by linear combinations of features.
- As a preprocessing step before classification algorithms like Logistic Regression, SVM, or KNN.

Steps in LDA:

4. Step 1: Compute the Mean Vectors - Calculate the mean vector for each class (e.g., red wine and white wine).
5. Step 2: Compute the Scatter Matrices - Compute within-class and between-class scatter matrices.
6. Step 3: Compute the Eigenvectors and Eigenvalues - Solve the generalized eigenvalue problem for $Sw^{-1}Sb$.
7. Step 4: Form the Linear Discriminants - Select the top eigenvectors to form the transformation matrix W.
8. Step 5: Project Data onto the New Subspace - Transform the original data to a lower-dimensional space using Y = XW.

Visualization:

After applying LDA, the transformed data can be visualized in 2D or 1D scatter plots to observe the separation between red and white wines.

Applications of LDA:

- Pattern recognition (e.g., face recognition using Fisherfaces).
- Medical diagnostics (classifying disease types).
- Marketing and customer segmentation.
- Financial analysis for credit scoring or risk classification.
- As a preprocessing step for classifiers.

Input:

The input for this experiment is the Wine dataset, which includes measurements for different wine characteristics and their respective types (class labels).

Output:

- A transformed dataset with reduced dimensions using LDA.
- Visualization showing clear class separation between red and white wines.

Conclusion:

Through the application of Linear Discriminant Analysis, we successfully transformed the high-dimensional wine data into a lower-dimensional space that highlights class separability. The experiment demonstrates how LDA can be used to enhance classification performance and visualize class distinctions effectively.

Outcome:

The outcome of this experiment is a low-dimensional representation of the wine dataset that maximizes class separability. This shows the effectiveness of LDA as a supervised feature transformation technique for classification tasks.

Questions:

1. What is the primary objective of applying the LDA algorithm in this experiment on the wine dataset? How does it differ from PCA in terms of its goal and output?

2. How does LDA achieve maximum class separability through the computation of scatter matrices?

3. Describe the step-by-step process of transforming the wine dataset using LDA. What are the mathematical operations involved in obtaining discriminant components?

4. How can the transformed data obtained from LDA be visualized to distinguish between red and white wines effectively?

5. Discuss the advantages and limitations of LDA in dimensionality reduction. How does the number of discriminant components affect the interpretability and classification accuracy of the model?