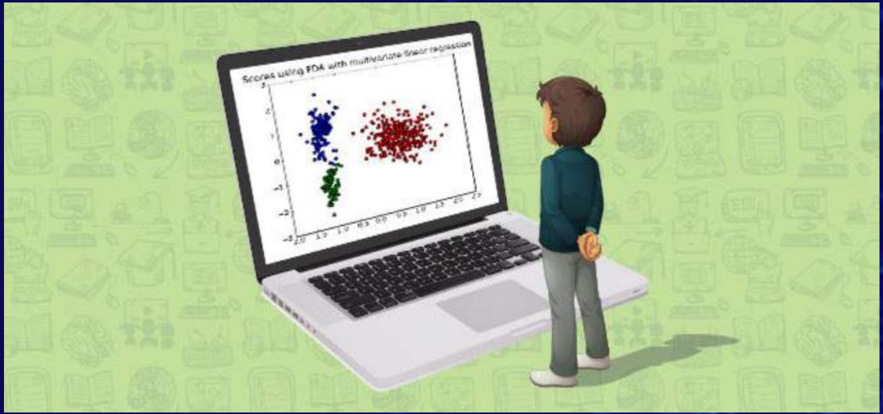
edunet foundation | Proudly supported by SAP | code unnati

Unit 3.5
Linear Discriminant Analysis in Machine Learning



Disclaimer: The content is curated for educational purposes only.

Disclaimer

The content is curated from online/offline resources and used for educational purpose only

Exam Score Analysis

Scenario: Imagine you are a teacher assessing the performance of your students in two subjects: Mathematics and English.

Data Collection: You have collected the scores of each student in both subjects for a class of 30 students.

Scenario Description: Upon plotting the scores of students in both subjects on a scatter plot, you notice that the data points are clustered in a way that suggests that the two subjects are correlated. However, it's not clear how well the scores distinguish high-performing students from low-performing ones.

Learning Objectives

- Introduction
- What is Linear Discriminant Analysis?
- Assumptions of Linear Discriminant Analysis
- How LDA works and Steps Involved?
- Hands On
- PCA vs LDA
- Applications of LDA
- Disadvantages of LDA
- Hands On

Introduction

- PCA aims to find the most accurate data representation in a lower dimensional space spanned by the maximum variance directions.
- However, such directions might not work well for tasks like classification.
- Here we present a new data reduction method that tries to preserve the discriminatory information between different classes of the dataset.

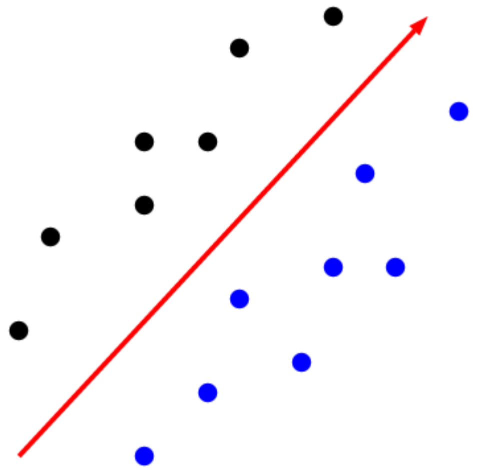


Image: Representative but not discriminative

What is Linear Discriminant Analysis (LDA)?

- Linear Discriminant Analysis (LDA) is most commonly used as dimensionality reduction technique in the pre-processing step for pattern-classification and machine learning applications.
- The goal is to project a dataset onto a lower-dimensional space with good class-separability in order avoid overfitting (“curse of dimensionality”) and reduce computational costs.
- LDA was developed as early as 1936 by Ronald A. Fisher.
- The original Linear discriminant applied to only a 2-class problem.
- It was only in 1948 that C.R. Rao generalized it to apply to multi-class problems.

What is Linear Discriminant Analysis (LDA).....?

- The main purpose of LDA is to find the line (or plane) that best separates data points belonging to different classes.
- The key idea behind LDA is that the decision boundary should be chosen such that it maximizes the distance between the means of the two classes.
- While maximizing distance LDA simultaneously minimizing the variance within each classes data or within-class scatter.
- This criterion is known as the Fisher criterion

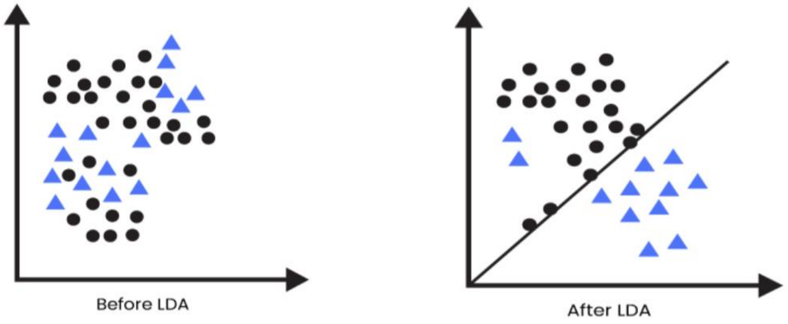


Image: LDA

Linear Discriminant Analysis Example

- For example: Consider a situation where you have plotted the relationship between two variables where each color represents a different class. One is shown with a red color and the other with blue in Image a.
- Suppose you want to reduce number of dimensions from 2 to 1, you can just project everything to the x-axis as shown below in Image b:

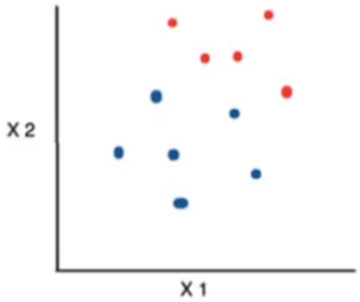


Image a

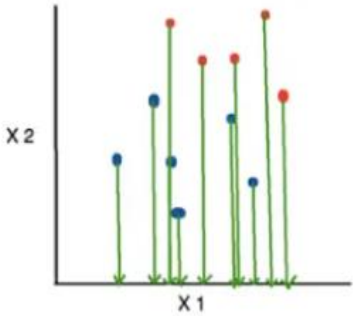
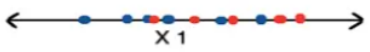


Image b



Result of projection to X axis

Reference : https://www.knowledgehut.com/blog/data-science/linear-discriminant-analysis-for-machine-learning

Linear Discriminant Analysis Example….

- This approach neglects any helpful information provided by the second feature. However, you can use LDA to plot it.
- The advantage of LDA is that it uses information from both the features to create a new axis which in turn minimizes the variance and maximizes the class distance of the two variables.

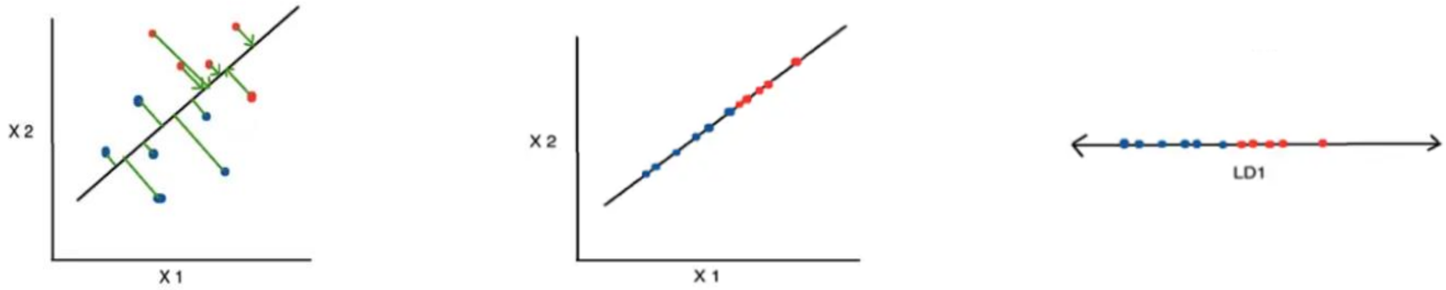


Image: LDA

Assumptions of Linear Discriminant Analysis

- Each feature (variable or dimension or attribute) in the dataset is a gaussian distribution. In other words, each feature in the dataset is shaped like a bell-shaped curve.
- Each feature has the same variance, the value of each feature varies around the mean with the same amount on average.
- Each feature is assumed to be randomly sampled.
- Lack of multicollinearity in independent features. Increase in correlations between independent features and the power of prediction decreases.

Linear Discriminant Analysis (LDA)

How LDA works and Steps involved?

1. Calculate the separability between different classes. This is also known as between-class variance and is defined as the distance between the mean of different classes.

"Between-class scatter matrix"

Overall mean

$$S_b = \sum_{i=1}^{g} N_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T$$

Sample size of class *i*

Sample mean of class *i*

Image: Between class variance

How LDA works and Steps involved?

2. Calculate the within-class variance. This is the distance between the mean and the sample of every class.

"Within-class scatter matrix"

Sample size

$$S_w = \sum_{i=1}^{g} (N_i - 1) S_i = \sum_{i=1}^{g} \sum_{j=1}^{N_i} (x_{i,j} - \bar{x}_i)(x_{i,j} - \bar{x}_i)^T$$

[Image: within class variance](#)

How LDA works and Steps involved?

3. Construct the lower-dimensional space that maximizes Step1 (between-class variance) and minimizes Step 2 (within-class variance). In the equation below P is the lower-dimensional space projection. This is also known as Fisher's criterion.

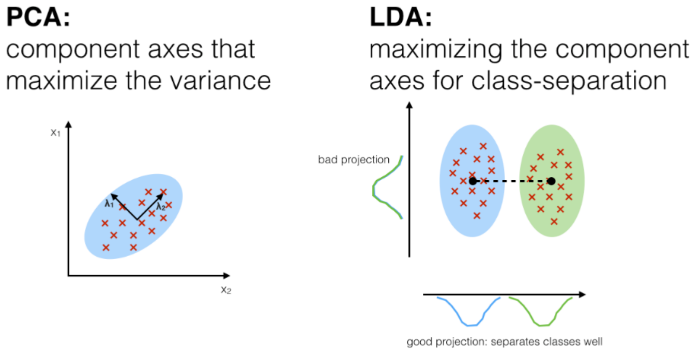
$$P_{lda} = \arg\max_{P} \frac{\left|P^T S_b P\right|}{\left|P^T S_w P\right|}$$

Image: Fisher's criterion

Lab 1 Implementation of LDA in Python using scikit-learn

Principal Component Analysis (PCA) vs. Linear Discriminant Analysis (LDA)

- PCA ignores class labels and focuses on finding the principal components that maximizes the variance in a given data. Thus it is an unsupervised algorithm.
- LDA is a supervised algorithm that intends to find the linear discriminants that represents those axes which maximize separation between different classes.
- LDA performs better multi-class classification tasks than PCA. However, PCA performs better when the sample size is comparatively small.



Applications of Linear Discriminant Analysis

- **Face Recognition:** LDA is used in face recognition to reduce the number of attributes to a more manageable number before the actual classification.
- **Medical:** You can use LDA to classify the patient disease as mild, moderate or severe. The classification is done upon the various parameters of the patient and his medical trajectory.
- **Customer Identification:** LDA helps in identifying and selecting which describes the properties of a group of customers who are most likely to buy a particular item in a shopping mall.
- **For predictions:** LDA is firmly used for prediction and hence in decision making, “will you read a book” gives you a predicted result through one or two possible class as a reading book or not.

Proudly supported by

Disadvantages of LDA

- LDA is used specifically in solving supervised classification problems for multiple classes; something impossible if using logistic regression. But LDA does not work in cases when the mean of the distributions is shared.
- In such a situation, LDA can not produce a new axis that can linearly separate both classes. To solve this problem, non-linear discriminant analysis is used in machine learning.
- One of the primary disadvantages of LDA is its sensitivity to the assumptions it relies on.
- LDA requires a sufficient number of data points compared to the number of features.

Lab 2 Comparison of PCA and LDA Using Wine Dataset

Proudly supported by

Summary

- Linear Discriminant Analysis (LDA) is a dimensionality reduction technique that focuses on enhancing class separability.
- It aims to maximize the distance between class means and minimize within-class variance.
- LDA is particularly useful for classification tasks where class distinctions are important.
- The steps of LDA involve computing class means, within-class scatter matrix, between-class scatter matrix, eigenvalue decomposition, and projection.
- LDA assumes normal distribution, equal covariance matrices, and independence of features within classes.
- Despite its advantages, LDA can be sensitive to small sample sizes and non-linear relationships.

Quiz

Question 1: Which step in LDA involves performing eigenvalue decomposition?

A) Calculating class means

B) Creating the projection matrix

C) Selecting top eigenvectors

D) Computing the between-class scatter matrix

Answer: B) Creating the projection matrix

Quiz

Question 2: LDA is most suitable for which type of data analysis tasks?

A) Clustering

B) Dimensionality reduction

C) Regression

D) Visualization

Answer: B) Dimensionality reduction

Quiz

Question 3: What is a key disadvantage of Linear Discriminant Analysis (LDA)?

A) Sensitivity to outliers

B) Inability to handle large datasets

C) No assumptions required

D) Works only for binary classification

Answer: A) Sensitivity to outliers

Quiz

Question 4: What is the primary goal of Linear Discriminant Analysis (LDA)?

A) Maximizing within-class variance

B) Minimizing between-class separation

C) Enhancing class separability

D) Reducing overall data variance

Answer: C) Enhancing class separability

Quiz

Question 5: In LDA, the within-class scatter matrix measures:

A) Separation between class means

B) Total variance within each class

C) Similarity between class centroids

D) Between-class distance

Answer: B) Total variance within each class

Thank you ...!