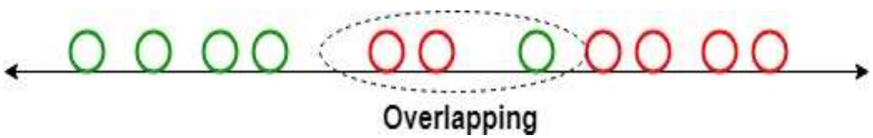
Linear Discriminant Analysis(LDA)

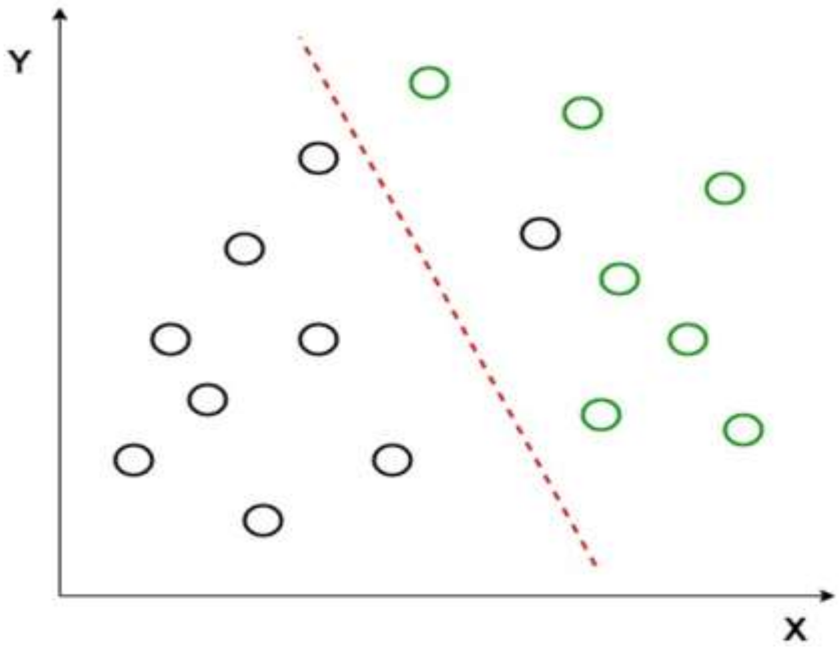
CONTENTS

What is LDA?

- Linear Discriminant Analysis (**LDA**) is a generalization of Fisher's linear discriminant, a method used in Statistics, pattern recognition and machine learning to find a linear combination of features that characterizes or separates two or more classes of objects or events.
- Linear Discriminant Analysis or LDA is a dimensionality reduction technique. It is used as a pre-processing step in Machine Learning and applications of pattern classification. The goal of LDA is to project the features in higher dimensional space onto a lower-dimensional space in order to avoid the curse of dimensionality and also reduce resources and dimensional costs.

Classification and Dimension Reduction

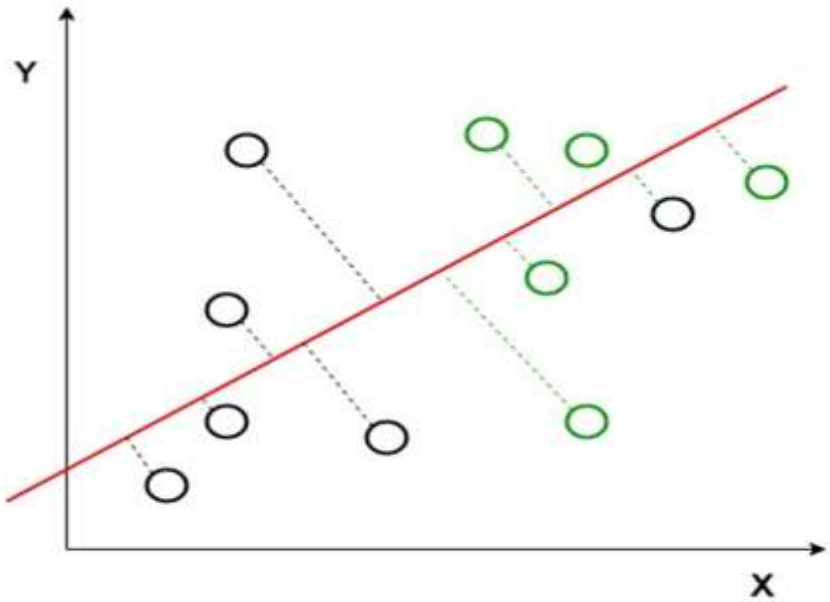




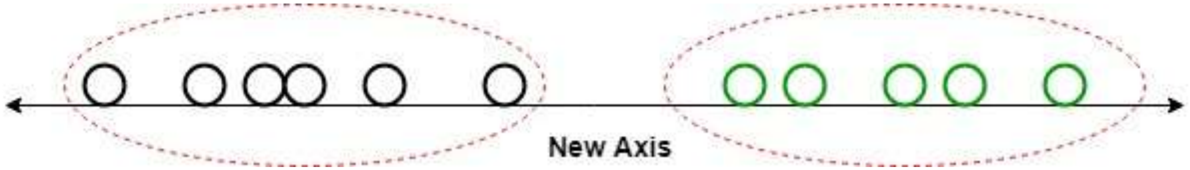
CLASSIFICATION AND DIMENSION REDUCTION

Two criteria are used by LDA to create a new axis:

1. Maximize the distance between means of the two classes.
2. Minimize the variation within each class.



CLASSIFICATION AND DIMENSION REDUCTION



HOW DOES LDA WORK?

LDA focuses primarily on projecting the features in higher dimension space to lower dimensions. You can achieve this in three steps:

Firstly, you need to calculate the separability between classes which is the distance between the mean of different classes. This is called the *between-class variance*.

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

- Secondly, calculate the distance between the mean and sample of each class. It is also called the within-class variance.

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

- Finally, construct the lower-dimensional space which maximizes the between-class variance and minimizes the within-class variance. P is considered as the lower-dimensional space projection, also called Fisher’s criterion.

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

Applications:

- **Face Recognition:** In the field of Computer Vision, face recognition is a very popular application in which each face is represented by a very large number of pixel values. Linear discriminant analysis (LDA) is used here to reduce the number of features to a more manageable number before the process of classification. Each of the new dimensions generated is a linear combination of pixel values, which form a template. The linear combinations obtained using Fisher's linear discriminant are called Fisher faces.
- **Medical:** In this field, Linear discriminant analysis (LDA) is used to classify the patient disease state as mild, moderate or severe based upon the patient various parameters and the medical treatment he is going through. This helps the doctors to intensify or reduce the pace of their treatment.
- **Customer Identification:** Suppose we want to identify the type of customers which are most likely to buy a particular product in a shopping mall. By doing a simple question and answers survey, we can gather all the features of the customers. Here, Linear discriminant analysis will help us to identify and select the features which can describe the characteristics of the group of customers that are most likely to buy that particular product in the shopping mall.

LDA VS. LR

(i) Two-Class vs. Multi-Class Problems

Logistic regression is both simple and powerful. However, it is traditionally used only in binary classification problems. While it can be extrapolated and used in multi-class classification problems, this is rarely done. When it’s a question of multi-class classification problems, linear discriminant analysis is usually the go-to choice. In fact, even with binary classification problems, both logistic regression and linear discriminant analysis are applied at times.

(ii) Instability With Well-Separated Classes

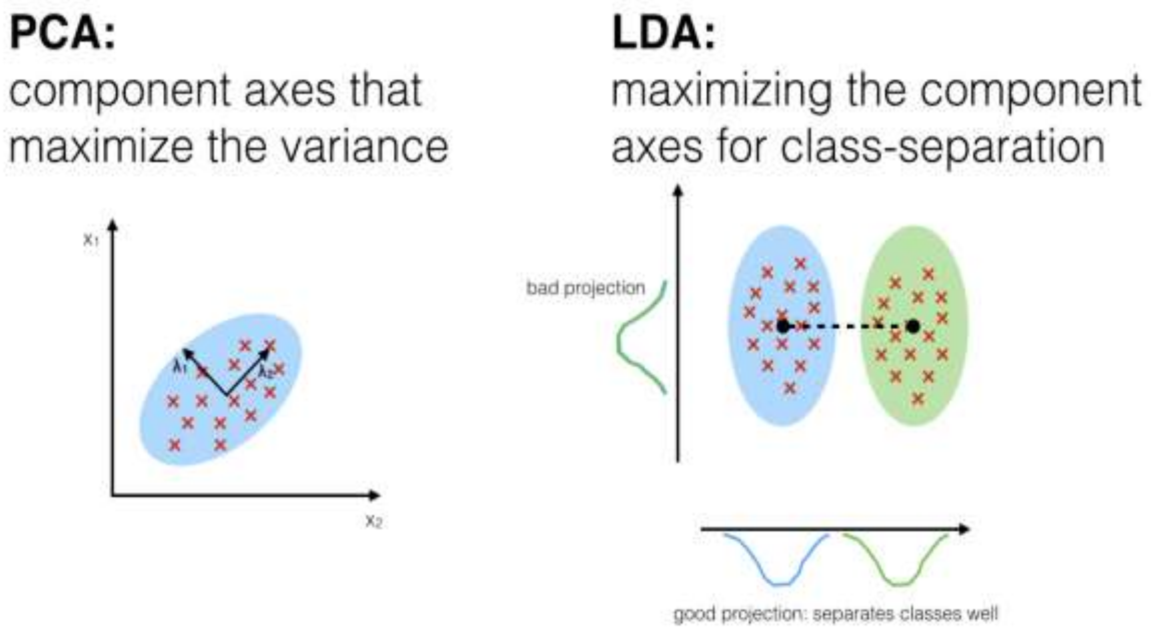
Logistic regression can become unstable when the classes are well-separated. This is where the Linear Discriminant Analysis comes in.

(iii) Instability With Few Examples

If there are just a few examples from the parameters need to be estimated, logistic regression tends to become unstable. In this situation too, Linear Discriminant Analysis is the superior option as it tends to stay stable even with fewer examples.

PCA vs. LDA

- Both **LDA and PCA** are linear transformation techniques: **LDA** is a supervised whereas **PCA** is unsupervised – **PCA** ignores class labels. In contrast to **PCA**, **LDA** attempts to find a feature subspace that maximizes class separability



A CONFUSION: IS LDA IS A LINEAR CLASSIFIER OR A DIMENSIONALITY REDUCTION TECHNIQUE?

The decision of LDA modes will depend of the dataset that we used, if it is linearly separable LDA as a classifier is a solution fast with great results, but if the characteristic of the dataset is nonlinear, LDA will be a extra tool to be applied over the dataset in order to try to “make things better” or facilitate the job of the posterior classifier.

Disadvantages

- It is not good for few dataset.
- It faces problems if data is very non-Gaussian.
- Sometimes not good for few categories variables
- Linear decision boundaries may not adequately separate the classes. Support for more general boundaries is desired.
- In a high-dimensional setting, LDA uses too many parameters. A regularized version of LDA is desired.
- Support for more complex prototype classification is desired.

Advantages

- Simple prototype classifier: Distance to the class mean is used, it's simple to interpret.
- The decision boundary is linear: It's simple to implement and the classification is robust.
- Dimension reduction: It provides an informative low-dimensional view on
  the data, which is both useful for visualization and feature engineering.
- 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.
- It still beats some algorithms (logistic regression) when its assumptions are met.

Thank You