SNS COLLEGE OF ENGINEERING

Kurumbapalayam (Po), Coimbatore – 641 107

An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A' Grade
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

COURSE NAME : 19CS501 Introduction to Machine Learning

III YEAR /V SEMESTER

Unit 2- SUPERVISED LEARNING

Topic : Discriminant Functions

Discriminant Analysis in Supervised Learning

- With the advancement in technology and trends in connected-devices could consider huge data into account, their storage and privacy is a big issue to concern

- Data hackers make algorithms to steal any such confidential information from a massive amount of data. So, data must be handled precisely which is also a time-consuming task.

- Also, we have seen, not all the data is required for inferences, reduction in data-dimensions can also help to govern datasets that could indirectly aid in the security and privacy of data.

Introduction to LDA

- In 1936, Ronald A.Fisher formulated Linear Discriminant first time and showed some practical uses as a classifier, it was described for a 2-class problem, and later generalized as 'Multi-class Linear Discriminant Analysis' or 'Multiple Discriminant Analysis' by C.R.Rao in the year 1948.

- Linear Discriminant Analysis is the most commonly used dimensionality reduction technique in supervised learning. Basically, it is a preprocessing step for pattern classification and machine learning applications.

- It projects the dataset into moderate dimensional-space with a genuine class of separable features that minimize overfitting and computational costs.

- With the aim to classify objects into one of two or more groups based on some set of parameters that describes objects, LDA has come up with specific functions and applications

Limitations of Logistic Regression

- **Logistic regression** is a simple and powerful linear classification algorithm. It also has limitations that suggest at the need for alternate linear classification algorithms.

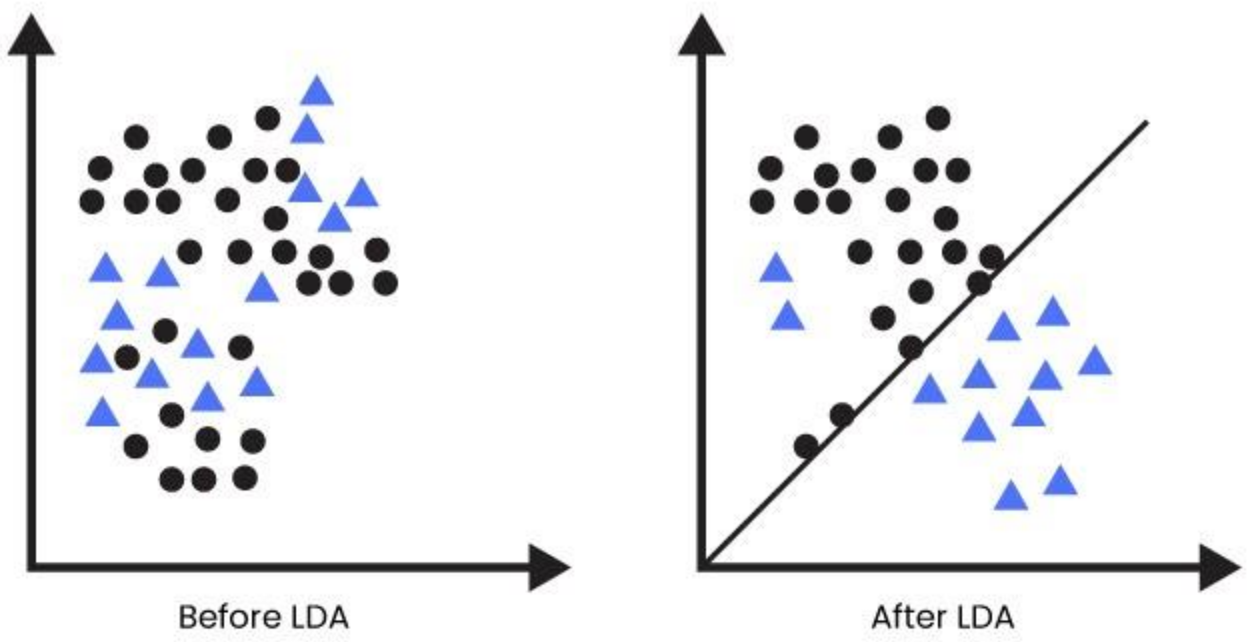
- **Two-Class Problems**. Logistic regression is intended for two-class or binary classification problems. It can be extended for multi-class classification, but is rarely used for this purpose.
- **Unstable With Well Separated Classes**. Logistic regression can become unstable when the classes are well separated.
- **Unstable With Few Examples**. Logistic regression can become unstable when there are few examples from which to estimate the parameters.
- **Linear Discriminant Analysis** does address each of these points and is the go-to linear method for multi-class classification problems. Even with binary-classification problems, it is a good idea to try both logistic regression and linear discriminant analysis.

Introduction to LDA

Which set of parameters can best describe the association of the group for an object?

What is the best classification preceptor model that separates those groups?

It is widely used for modeling varieties in groups, i.e. distributing variables into two or more classes, suppose we have two classes and we need to classify them efficiently.



Classes can have multiple features, using one single feature to classify may yield in some kind of overlapping of variables, so there is a need of increasing the number of features to avoid overlapping that would result in proper classification in return.

Example of LDA

Consider another simple example of dimensionality reduction and feature extraction, you want to check the quality of soap based on the information provided related to a soap including various features such as weight and volume of soap, peoples' preferential score, odor, color, contrasts, etc.

A small scenario to understand the problem more clearly;

Object to be tested -Soap;

To check the quality of a product- class category as 'good' or 'bad'(dependent variable, categorical variable, measurement scale as a nominal scale);

Features to describe the product- various parameters that describe the soap (independent variable, measurement scale as nominal, ordinal, internal scale);

Example of LDA

When the target variable or dependent variable is decided then other related information can be dragged out from existing datasets to check the effectivity of features on the target variables.

And hence, the data dimension gets reduced out and important related-features have stayed in the new dataset.

How to Prepare Data for LDA

- **Classification Problems**. This might go without saying, but LDA is intended for classification problems where the output variable is categorical. LDA supports both binary and multi-class classification.

- **Gaussian Distribution**. The standard implementation of the model assumes a Gaussian distribution of the input variables. Consider reviewing the univariate distributions of each attribute and using transforms to make them more Gaussian-looking (e.g. log and root for exponential distributions and Box-Cox for skewed distributions).

- **Remove Outliers.** Consider removing outliers from your data. These can skew the basic statistics used to separate classes in LDA such the mean and the standard deviation.

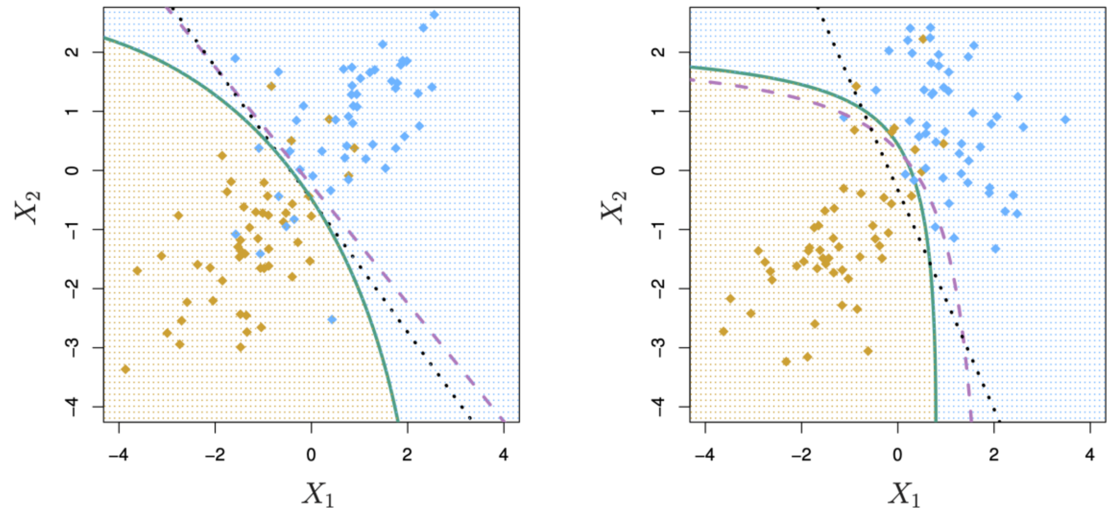
- **Same Variance**. LDA assumes that each input variable has the same variance. It is almost always a good idea to standardize your data before using LDA so that it has a mean of 0 and a standard deviation of 1.

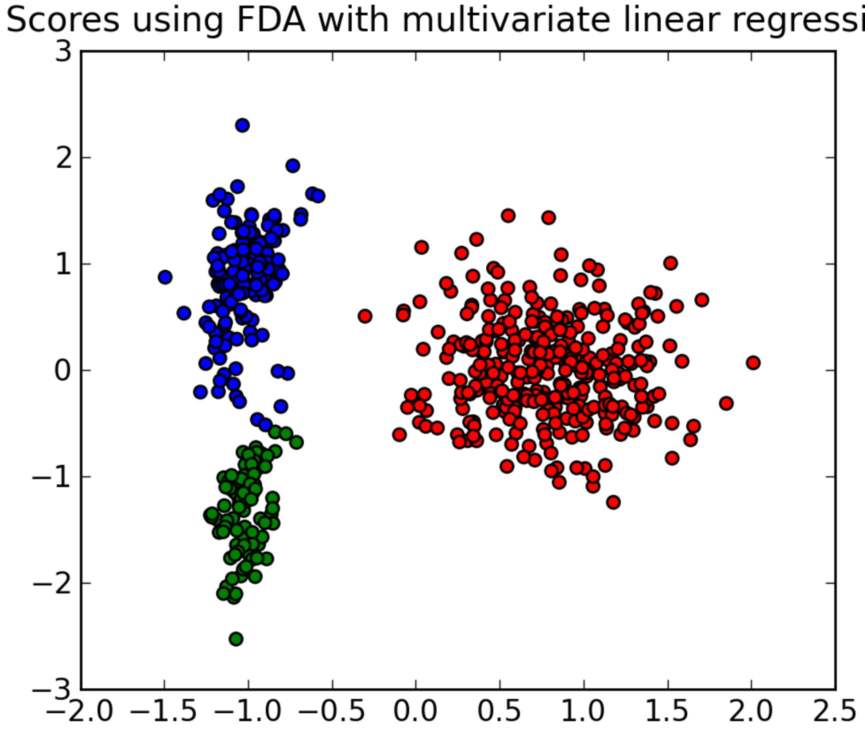
Extensions to LDA

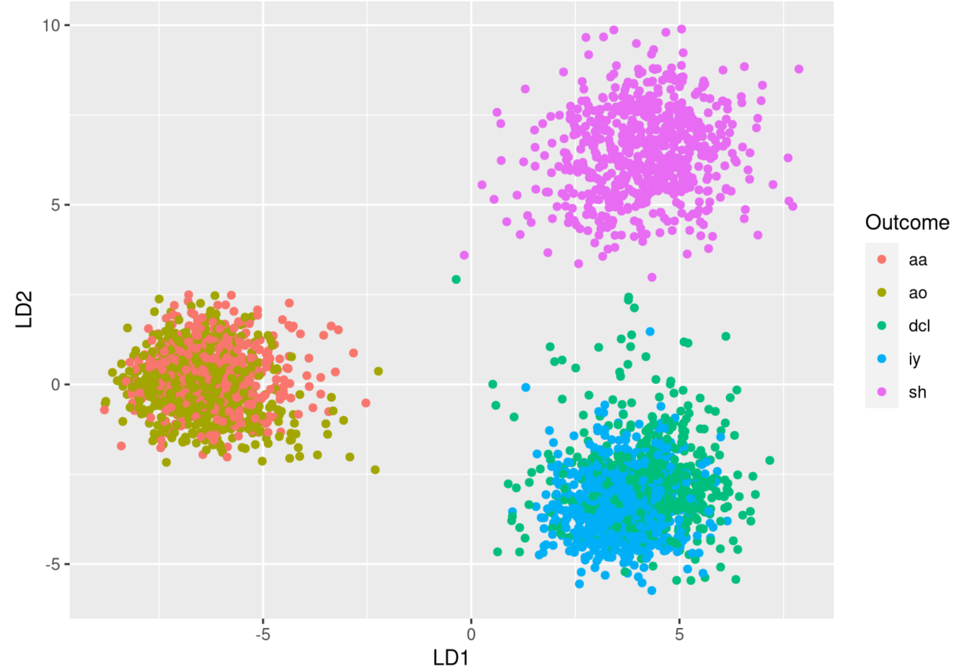
- Quadratic Discriminant Analysis (QDA): Each class deploys its own estimate of variance, or the covariance where there are multiple input variables.

- Flexible Discriminant Analysis (FDA): Where the combinations of non-linear sets of inputs are deployed such as splines.

- Regularized Discriminant Analysis (RDA): It adds regularization into the estimate of the variance, or covariance that controls the impact of various variables on LDA. (Source)







Limitations of Logistic Regression

- **Two-Class Problems**: Logistic regression is proposed for two-class or binary classification problems that further be expanded for multi-class classification, but is rarely used for this purpose.

- **Unstable With Well Separated Classes: Logistic regression is restricted and unstable when the classes are well-separated.**

- **Unstable With Few Examples:** Logistic regression behaves as an unstable method while dealing with few examples from which parameters are estimated

Linear Discriminant Analysis can handle all the above points and acts as the linear method for multi-class classification problems

Working of Linear Discriminant Analysis

- **Assumptions**

1. Every feature either be variable, dimension, or attribute in the dataset has gaussian distribution, i.e, features have a bell-shaped curve.

2. Each feature holds the same variance, and has varying values around the mean with the same amount on average.

3. Each feature is assumed to be sampled randomly.

4. Lack of multicollinearity in independent features and there is an increment in correlations between independent features and the power of prediction decreases.

Working of Linear Discriminant Analysis

- While focusing on projecting the features in higher dimension space onto a lower dimensional space, LDA achieve this via three step process;

- **First step**: To compute the separate ability amid various classes,i.e, the distance between the mean of different classes, that is also known as between-class variance.

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

Working of Linear Discriminant Analysis

- **Second Step**: To compute the distance among the mean and sample of each class,that is also known as the within class variance.

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

- **Third step**: To create the lower dimensional space that maximizes the between class variance and minimizes the within class variance.

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

Application of Linear Discriminant Analysis

- There are various techniques used for the classification of data and reduction in dimension, among which Principal Component Analysis(PCA) and Linear Discriminant Analysis(LDA) are commonly used techniques.

- The condition where within -class frequencies are not equal, Linear Discriminant Analysis can assist data easily, their performance ability can be checked on randomly distributed test data.

- This method results in the maximization of the ratio between-class variance to the within-class variance for any dataset and maximizes separability.

- LDA has been successfully used in various applications, as far as a problem is transformed into a classification problem, this technique can be implemented.

Application of Linear Discriminant Analysis

- For example, LDA can be used as a classification task for speech recognition, microarray data classification, face recognition, image retrieval, bioinformatics, biometrics, chemistry, etc. below are other applications of LDA;

- For customers' recognition: LDA helps here to identify and choose the parameters to describe the components of a group of customers who are highly likely to buy similar products.

- For face recognition: it is the most famous application in the field of computer vision, every face is drawn with a large number of pixel values. Here, LDA reduces the number of features to a more controllable number first before implementing the classification task. A temple is created with newly produced dimensions which are a linear combination of pixel values.

- In medical: LDA is used here to classify the state of patients' diseases as mild, moderate or severe based on the various parameters and the medical treatment the patient is going through in order to decrease the movement of treatment.

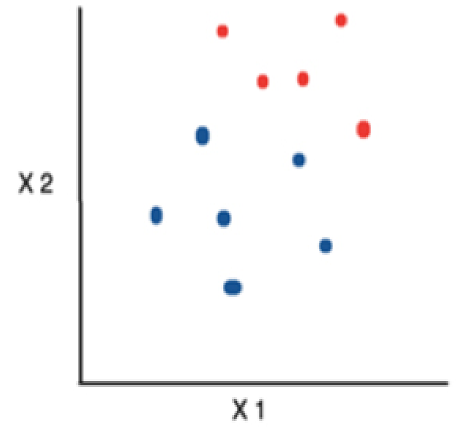
Application of Linear Discriminant Analysis

- 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.

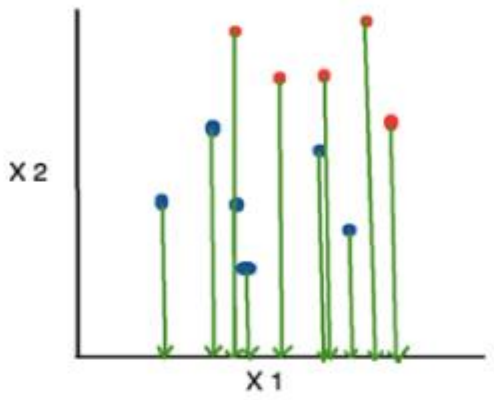
- In learning: Nowadays, robots are trained to learn and talk to work as human beings, this can be treated as classification problems. LDA makes similar groups based on various parameters such as frequencies, pitches, sounds, tunes, etc.

Application of Linear Discriminant Analysis

- 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

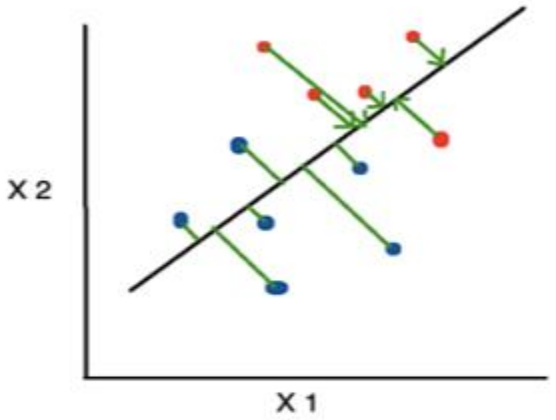


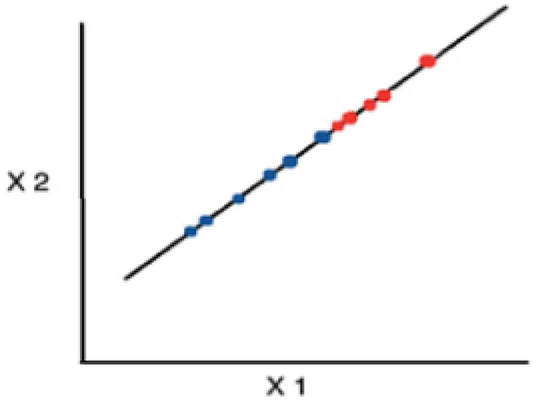
If you are willing to reduce the number of dimensions to 1, you can just project everything to the x-axis as shown below:



Application of Linear Discriminant Analysis

- 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.



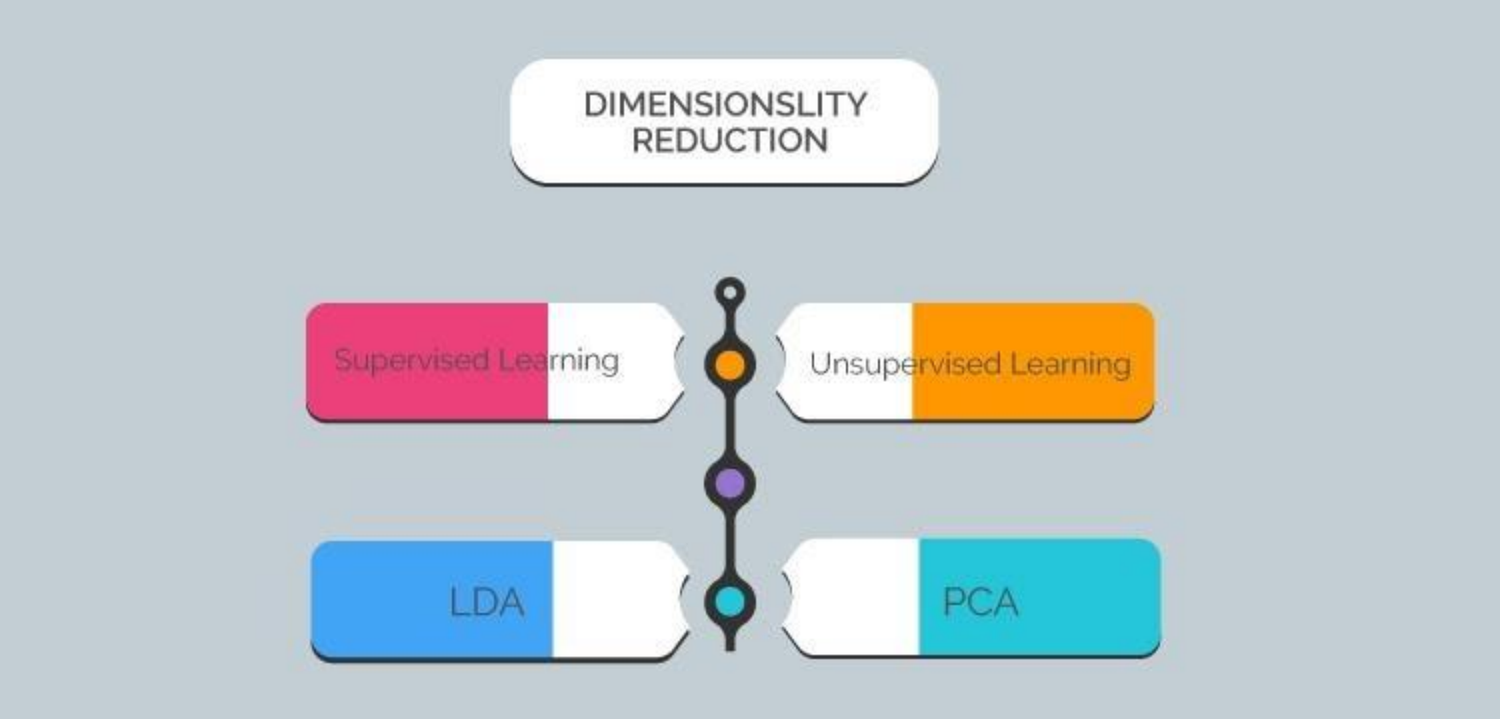


LDA vs PCA

- ➢ From the above discussion, we came to know that in general, the LDA approach is very similar to Principal Component Analysis, both are linear transformation techniques for dimensionality reduction, but also pursuing some differences;

- ❖ The earliest difference between LDA and PCA is that PCA can do more of features classification and LDA can do data classification.

- ❖ The shape and location of a real dataset change when transformed into another space under PCA, whereas
- ➢ PCA can be expressed as an unsupervised algorithm since it avoids the class labels and focuses on finding directions(principal components) to maximize the variance in the dataset,

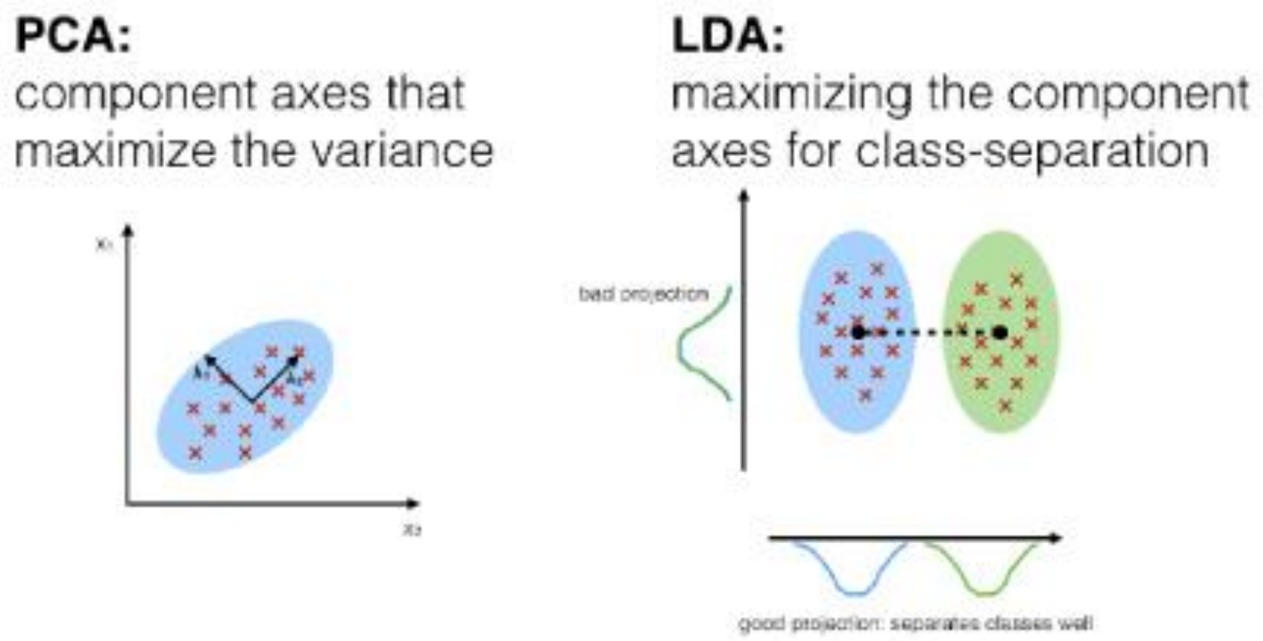


LDA vs PCA

- 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. On the other hand, LDA is a supervised algorithm that intends to find the linear discriminants that represents those axes which maximize separation between different classes.

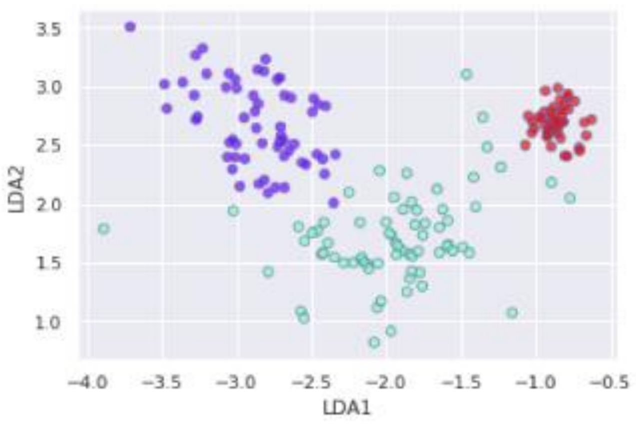
- LDA performs better multi-class classification tasks than PCA. However, PCA performs better when the sample size is comparatively small. An example would be comparisons between classification accuracies that are used in image classification.

- Both LDA and PCA are used in case of dimensionality reduction. PCA is first followed by LDA.



Assessment

➢ How to implement LDA using scikit-learn?



REFERENCES

1. Tom M. Mitchell, “Machine Learning”, McGraw-Hill Education (India) Private Limited, 2013.
2. Trevor Hastie, Robert Tibshirani, Jerome Friedman, “The Elements of Statistical Learning: Data Mining, Inference, and Prediction”, Springer; Second Edition, 2009.

THANK YOU