VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

SCHOOL OF INFORMATION TECHNOLOGY AND ENGINEERING

ITA5007-Data Mining and Business Intelligence

MCA

Topic :

1. Linear Discriminant Analysis
2. Example (implementation of LDA on Iris dataset)

Description:

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 to avoid overfitting ("curse of dimensionality") and also reduce computational costs.

Ronald A. Fisher formulated the Linear Discriminant in 1936.

The original Linear discriminant was described for a 2-class problem, and it was then later generalized as "multi-class Linear Discriminant Analysis" or "Multiple Discriminant Analysis" by C. R. Rao in 1948.

The general LDA approach is very similar to a Principal Component Analysis, but in addition to finding the component axes that maximize the variance of our data (PCA), there is additional interest in the axes that maximize the separation between multiple classes (LDA).

So, in a nutshell, often the goal of LDA is to project a feature space (a dataset of n-dimensional samples) onto a smaller subspace k (where $k \leq n-1$) while maintaining the class-discriminatory information.

In general, dimensionality reduction does not only help reduce computational costs for a given classification task, but it can also be helpful to avoid overfitting by minimizing the error in parameter estimation.

Linear Discriminant Analysis or Normal Discriminant Analysis or Discriminant Function analysis is a dimensionality reduction technique commonly used for classification problems.

For example, we have two classes and we need to separate them efficiently. Classes can have multiple features. Using only a single feature to classify them may result in some overlapping. So, we will keep on increasing the number of features for proper classification. From the n independent variables of the dataset, LDA extracts p<=n new independent variables that separate the most classes of the dependent variable. The fact that dependent variable is considered makes the LDA a supervised model.

LDA can be achieved in three steps:

1. The first step is to calculate the severability between different classes (i.e. the distance between the mean of different classes) also called between class variance.
2. The second step is to calculate the distance between the mean and sample of each class, which is called the within class variance.
3. The third step is to construct the lower dimensional space which maximizes between class variance and minimizes the within class variance.

VIT®

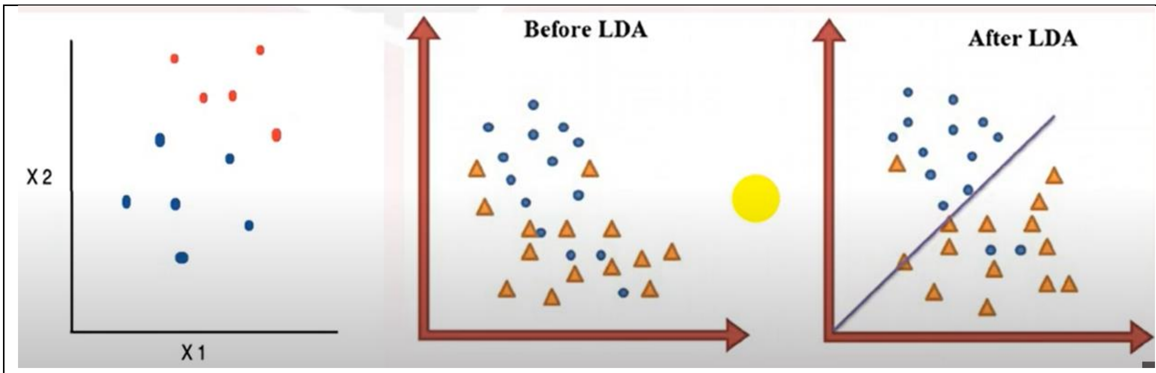
Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

SCHOOL OF INFORMATION TECHNOLOGY AND ENGINEERING

ITA5007-Data Mining and Business Intelligence

MCA



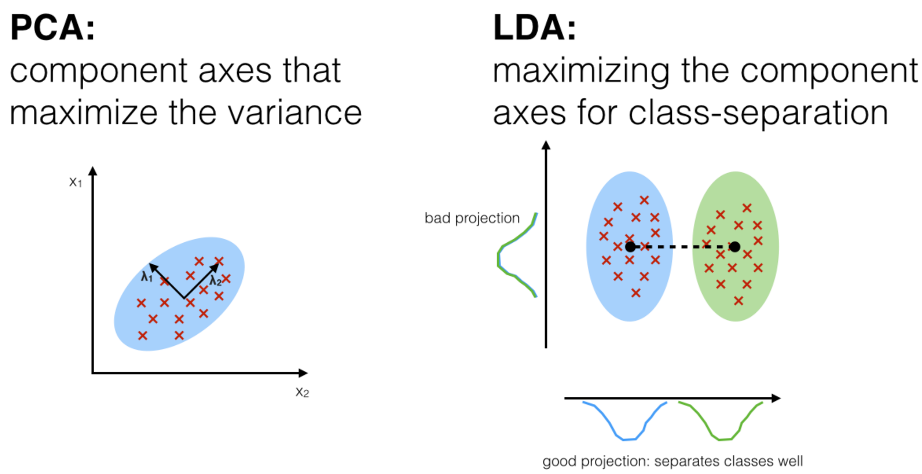
Principal Component Analysis vs. Linear Discriminant Analysis

Both Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA) are linear transformation techniques that are commonly used for dimensionality reduction.

PCA can be described as an “unsupervised” algorithm, since it “ignores” class labels and its goal is to find the directions (the so-called principal components) that maximize the variance in a dataset.

In contrast to PCA, LDA is “supervised” and computes the directions (“linear discriminants”) that will represent the axes that that maximize the separation between multiple classes.

In practice, it is also not uncommon to use both LDA and PCA in combination: E.g., PCA for dimensionality reduction followed by an LDA.



Summarizing the LDA approach in 5 steps

Listed below are the 5 general steps for performing a linear discriminant analysis;

1. Compute the d-dimensional mean vectors for the different classes from the dataset.
2. Compute the scatter matrices (in-between-class and within-class scatter matrix).
3. Compute the eigenvectors (e_1, e_2, ..., e_d) and corresponding eigenvalues ($\lambda_1,\lambda_2,...,\lambda_d$) for the scatter matrices.

4. Sort the eigenvectors by decreasing eigenvalues and choose k eigenvectors with the largest eigenvalues to form a d×k dimensional matrix W (where every column represents an eigenvector).
5. Use this d×k eigenvector matrix to transform the samples onto the new subspace.
   This can be summarized by the matrix multiplication: Y=X×W (where X is a n×d-dimensional matrix representing the n samples, and y are the transformed n×k-dimensional samples in the new subspace).

Courtesy : https://sebastianraschka.com/Articles/2014_python_lda.html

SCHOOL OF INFORMATION TECHNOLOGY AND ENGINEERING

ITA5007-Data Mining and Business Intelligence

MCA