Graphical user interface, text, application

Description automatically generatedGraphical user interface, application

Description automatically generated

Student Name

***Sanatkumar Rajmogali Ippalpalli***

Title of Project Report

***Guided Project 2 – Linear Discriminant Analysis***

Table of Contents

[Introduction 4](#_Toc82526209)

[Eckovation theme & Question 5](#_Toc82526210)

[Prerequisites before starting coding 6](#_Toc82526211)

[program DEVELOPMENT steps 8](#_Toc82526212)

[Dataset/Image requirements 8](#_Toc82526213)

[Technique – LINEAR dISCRIMINANT ANALYSIS 8](#_Toc82526214)

[PROGRAM / CODE DEVELOPMENT 9](#_Toc82526215)

[Analysis 14](#_Toc82526216)

[CONCLUSION 15](#_Toc82526217)

[Figure 1 Import libraries and datasets/modules 9](#_Toc82526199)

[Figure 2 Load Pandas Dataset 10](#_Toc82526200)

[Figure 3 Visualization of Dataset 11](#_Toc82526201)

[Figure 4 Andrews curves 11](#_Toc82526202)

[Figure 5 dataset iloc code 12](#_Toc82526203)

[Figure 6 Linear Discriminant Analysis on Dataset 12](#_Toc82526204)

[Figure 7 Accuracy Score 13](#_Toc82526205)

[Figure 8 Classification 13](#_Toc82526206)

[Figure 9 Confusion matrix 13](#_Toc82526207)

[Figure 10 Predict & Accuracy 14](#_Toc82526208)

***EXECUTIVE SUMMARY***

Science and technology improved many technologies and has guided numerous innovative features which advanced dimensionality reduction techniques.

As part of guided project, one of the important dimensionality reduction technology is **Linear Discriminant Analysis**  used for the supervised classification problems.

Linear Discriminant Analysis is a well-known scheme for

feature extraction and dimension reduction

Linear Discriminant Analysis is a well-known scheme for

feature extraction and dimension reduction

Linear Discriminant Analysis is a well-known scheme for

feature extraction and dimension reduction

Linear Discriminant Analysis is a well-known scheme for feature extraction and dimension reduction. It has been used widely in many applications such as face recognition, image retrieval, microarray data classification, etc. Classical LDA projects the data onto a lower-dimensional vector space such that the ratio of the between-class distance to the within-class distance is maximized, thus achieving maximum discrimination.

It is used for modelling differences in groups i.e. separating two or more classes. It is used to project the features in higher dimension space into a lower dimension space.

# Introduction

Science and technology improved many technologies and has guided numerous innovative features which advanced the techniques in data classification.

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 also reduce computational costs.

**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), we are additionally interested in the axes that maximize the separation between multiple classes (LDA).**

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

Hence Eckovation includes this guided project in the courseware for students to understand, implementation / execute the code themselves.

This report includes the 5W1H about the theme of development of code and running the code with database available over the internet. At the end of the report, the conclusions share the adaptive thresholding & OTSU thresholding features extracted and useful for next course of activities to gain advantages in the edge detection activities development.

# Eckovation theme & Question

**Theme : Linear Discriminant Analysis**

**Linear Discriminant Analysis or LDA** is a dimensionality reduction technique used to reduce the number of dimensions (i.e. variables) in a dataset while retaining as much information as possible. Using LDA based classification, we can find discriminative features for a given audio segment to achieve the task of Automatic Speech Classification such that speech belonging to the same class are close together, but samples from different classes are far apart from each other.

**Question:**

In order to implement LDA, first generate a dummy dataset (say **IRIS dataset having 4 features**) and the use LDA to decrease the number of **features to one/two**. Now using this modified dataset, try to learn a classifier to test the **performance of LDA** for dimensionality reduction.

# Prerequisites before starting coding

1. Who - Software needed?
2. What - Version / Release of software?
3. Any Prerequisites
4. How - to install the software
5. Which -libraries are needed to execute the problem statement
6. Where – dataset requirements, path location to include in the code
7. When – to use the above feature extraction
8. Who – Software neeed?

Python

1. What- Version / Release of software?

Python version 3.6 (latest version of python)

1. Any Prerequisites

RAM space availability & hard disk space availability

Admin rights to install the software

1. How - to install the software
2. The following url <https://www.python.org/downloads/>can be referred to download python.
3. Second and easier option is to download anaconda and use its anaconda prompt to run the commands. To install anaconda check this url <https://www.anaconda.com/download/>
4. Which -libraries are needed to execute the problem statement
5. Sklearn dataset
6. Numpy (pip install numpy)
7. Matplotlib (pip install matplotlib)
8. Where – dataset requirements, path location to include in the code
9. Once you have python downloaded and installed, you will need to setup PATH variables (if you want to run python program directly, detail instructions are below in how to run software section). To do that check this: [https://www.pythoncentral.io/add-python-to-path-python-is-not- recognized-as-an-internal-or-external-](https://www.pythoncentral.io/add-python-to-path-python-is-not-recognized-as-an-internal-or-external-command/) [command/](https://www.pythoncentral.io/add-python-to-path-python-is-not-recognized-as-an-internal-or-external-command/).
10. Setting up PATH variable is optional as you can also run program without it and more instruction are given below on this topic.
11. When – to use the above feature extraction
12. When – to use the above technique

It is used for modelling differences in groups i.e. separating two or more classes. It is used to project the features in higher dimension space into a lower dimension space. Dimensionality reduction does not only help reducing computational costs for a given classification task.

# program DEVELOPMENT steps

* Dataset/Image requirement
* Technique selections
* Program / code development
* Analysis

### Dataset/Image requirements

The image sources used for this project are downloaded / collected from internet.

Purple Flower. JPG – significant flower image with color may help to find edges

Maheshbabu.JPG – Famous south Indian Tollywood actor

Allfruits.jpg – All color combinations

### Technique – LINEAR dISCRIMINANT ANALYSIS

For performing a linear discriminant analysis, following simple steps to be implemented.

1. Compute the dd-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 (ee1,ee2,...,eedee1,ee2,...,eed) and corresponding eigenvalues (λλ1,λλ2,...,λλdλλ1,λλ2,...,λλd) for the scatter matrices.
4. Sort the eigenvectors by decreasing eigenvalues and choose kk eigenvectors with the largest eigenvalues to form a d×kd×k dimensional matrix WWWW (where every column represents an eigenvector).
5. Use this d×kd×k eigenvector matrix to transform the samples onto the new subspace. This can be summarized by the matrix multiplication: YY=XX×WWYY=XX×WW (where XXXX is a n×dn×d-dimensional matrix representing the nn samples, and yyyy are the transformed n×kn×k-dimensional samples in the new subspace).

Let us hop to the inscribing carving!

### PROGRAM / CODE DEVELOPMENT

As explained step by step during the lecture by mentor, we would approach steps and understand the basics with brief explanation as needed.

#### Step 1: Import the relevant libraries and applicable datasets/modules

Graphical user interface, text, application

Description automatically generated

Figure 1 Import libraries and datasets/modules

#### Step 2: Load dataset/Image

As Eckovation advised to use famous “Iris” dataset that has been deposited on the UCI machine learning repository (https://archive.ics.uci.edu/ml/datasets/Iris).

The iris dataset contains measurements for 150 iris flowers from three different species.A bee on a purple flower

Description automatically generatedThe three classes in the Iris dataset:

1. Iris-setosa (n=50)
2. Iris-versicolor (n=50)
3. Iris-virginica (n=50)

The four features of the Iris dataset:

1. sepal length in cm
2. sepal width in cm
3. petal length in cm
4. petal width in cm

Table

Description automatically generated

Figure 2 Load Pandas Dataset

Features extraction & Visulization of Dataset

Text, table

Description automatically generated

Table

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Chart, scatter chart

Description automatically generated

Figure 3 Visualization of Dataset

#### Step 3: Andrews curves

Chart

Description automatically generated

Figure 4 Andrews curves

Graphical user interface, text

Description automatically generated

Figure 5 dataset iloc code

#### Step 4: Linear Discriminant Analysis

Apply the linear discriminant analysis technique

Graphical user interface, text, application, email

Description automatically generated

Figure 6 Linear Discriminant Analysis on Dataset

Graphical user interface, text, application

Description automatically generated

Figure 7 Accuracy Score

#### Step 5: CLASSIFICATION & CONFUSION MATRIX

Table

Description automatically generated

Figure 8 Classification

Graphical user interface, text, application

Description automatically generated

Figure 9 Confusion matrix

Graphical user interface, text, application, email

Description automatically generated

Figure 10 Predict & Accuracy

### Analysis

The LDA is completed on IRIS dataset featuring different classes with four features as sepal length, sepal width, petal length, petal width.

This entire program runs within few seconds.

# CONCLUSION

In this guided project, we built linear discriminant analysis technique to reduce number of dimension (class).

This is done in first attempt with perfect score 1.0. However, the improvements in the code with time with multiple attempts may be checked for various comparisons with other dimensionality techniques & and justified for the accuracy score.

This entire program runs within few seconds.

references:

1. <https://sebastianraschka.com/Articles/2014_python_lda.html>
2. <https://www.geeksforgeeks.org/ml-linear-discriminant-analysis/>