GENG-8900 MACHINE LEARNING

Instructor: Dr. Yasser Alginahi

Dimensionality reduction: Linear Discriminant Analysis

Shashank Prakash Naidu	110090086
Tejus Vivek	110095845
Mahenur Patel	110120170

Date: 17-Nov-2023

What is Dimensionality Reduction?

- Dimensionality reduction is a technique used to reduce the number of features in a dataset while retaining as much of the important information as possible.[1]

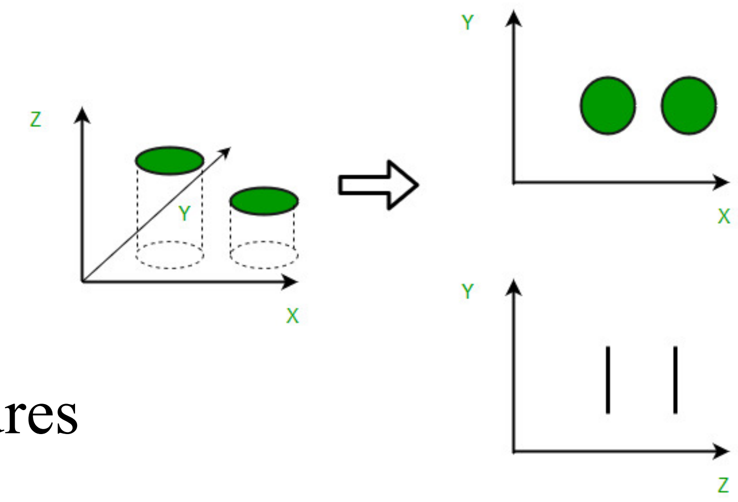


Fig.1 Components of Dimensionality Reduction [https://tinyurl.com/5n6k7r2e]

High- Dimensional data ⟶ Large Number of Variables / Features

There are two main approaches to dimensionality reduction:

- Feature selection
- Feature extraction.

Why Dimensionality Reduction?

- High-dimensional data poses challenges in analysis and interpretation.
- Dimensionality reduction enhances computational efficiency and aids in visualization.
- We'll explore the motivations behind the need for techniques like LDA.

Example: Think about a dataset tracking customer behavior on an e-commerce platform.

- **Original Dataset:** customers' age, browsing history, purchase frequency, and more.

- **After Dimensionality Reduction:** essential patterns in customer behavior.

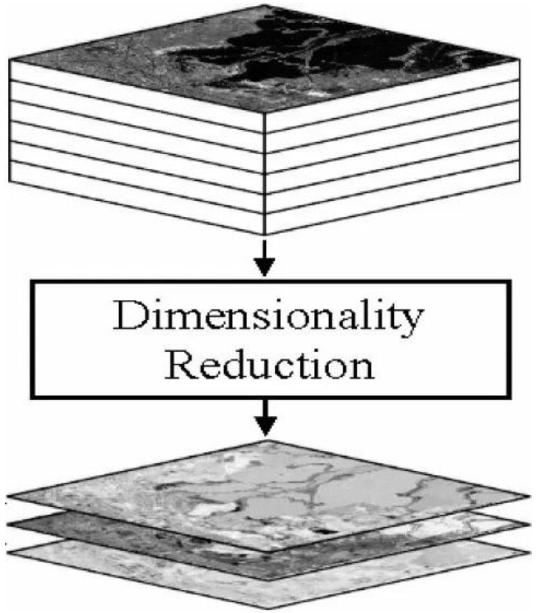


Fig.2 Dimensionality Reduction
[https://tinyurl.com/3mx28kj7]

What is Linear Discriminant Analysis?

- Linear Discriminant Analysis (LDA) is a supervised learning algorithm used for classification tasks in machine learning. It is a technique used to find a linear combination of features that best separates the classes in a dataset.[2]

Higher Dimensional Space(Data) ⟶ LDA ⟶ Lower Dimensional Space(Data)

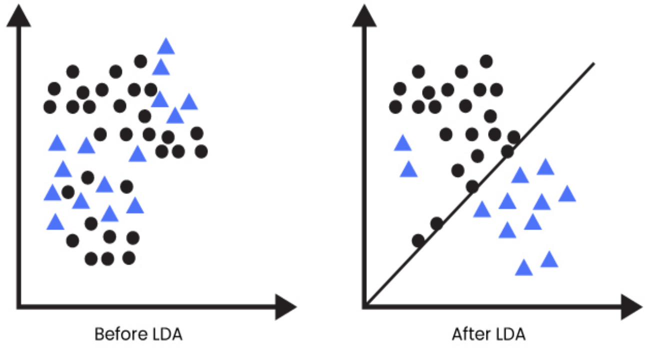


Fig3. Classification of various objects before and after implementing LDA (https://tinyurl.com/23ywuc56)

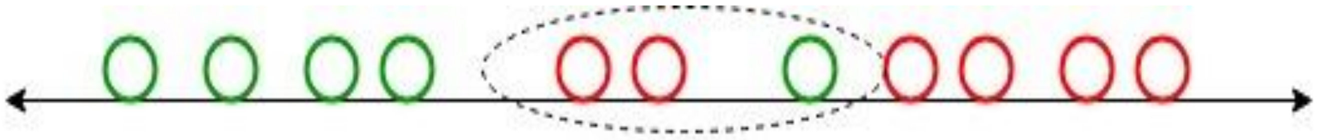
Two criteria are used by LDA to create a new axis (Lower Dimensional Space):

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

Overlapping in Linear Discriminant Analysis (LDA)

Example: Apple and Tomato

- In this case, there are two classes, and we must effectively divide them. Classes can have several features. As the image below illustrates, there may be some overlap if you try to categorize them using only one feature.



Overlapping (https://tinyurl.com/66snrywh)

2.4mg	Fiber	1.2gm
54 IU	Vitamin A	833 IU
5mg	Magnesium	11mg

Table1.Features of Apple and Tomato
(https://tinyurl.com/4vp5b7s6)

Better Understanding in Linear Discriminant Analysis (LDA)

- Two sets of data points belonging to two different classes.
- Data points are plotted on 2D plane.
- No straight line can be seen.
- LDA uses both X and Y axis to create a new axis.
- LDA is used to reduce to 2D graph to 1D graph.

○ ⟶ Data set Apple

○ ⟶ Data set B Tomato

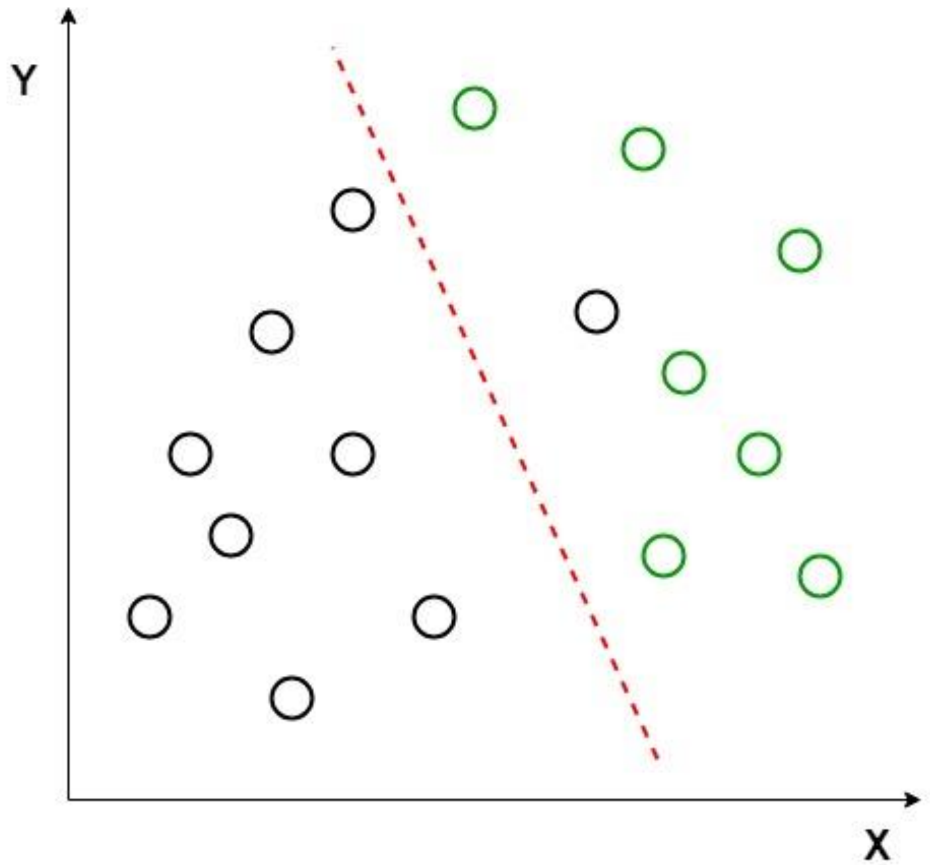


Fig4. Representation of two sets of data points(https://tinyurl.com/66snrywh)

Better Understanding in Linear Discriminant Analysis (LDA) (Contd…)

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

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

- Red line indicates the new axis that has been generated.
- The generated axis increases the separation between the data points of the two classes.

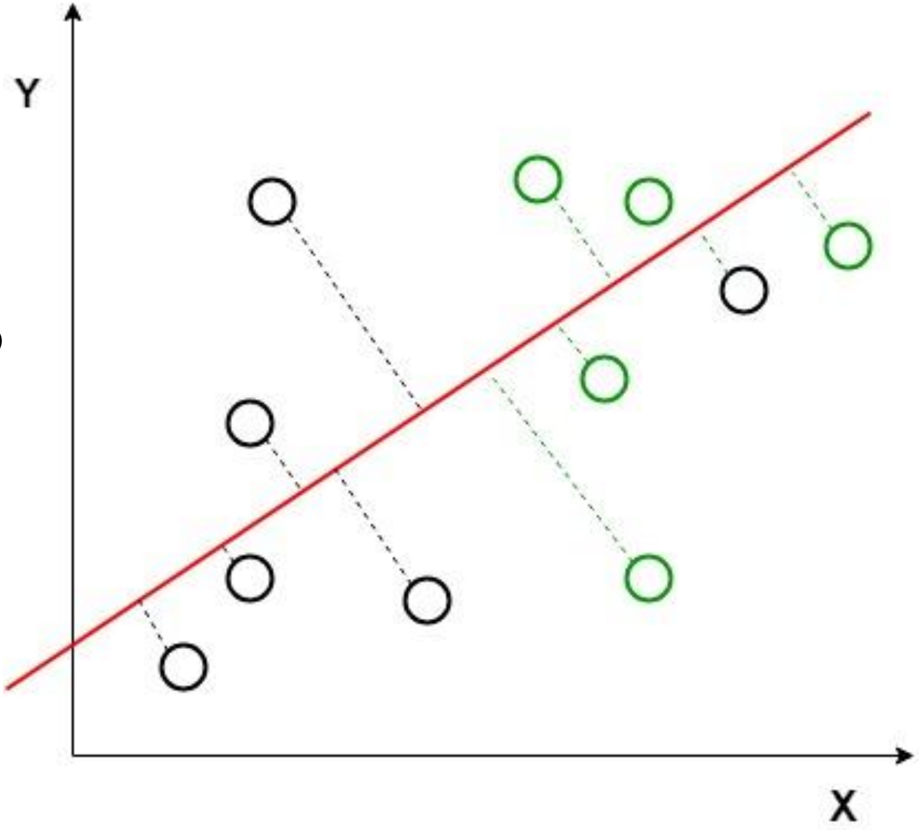


Fig5. Data points after LDA
(https://tinyurl.com/66snrywh)

Better Understanding in Linear Discriminant Analysis (LDA) (Contd…)

- All the data points of the classes are plotted on this new axis and are shown in the fig6. given below. .

Fig6. Data points on 1D graph after LDA(https://tinyurl.com/66snrywh)

- When the distribution means are similar, LDA is unable to identify a new axis that would allow both classes to be linearly separable, so the analysis is deemed ineffective. We apply non-linear discriminant analysis in these situations.

PCA vs LDA?

- Principal Component Analysis is
  an unsupervised learning technique [1]
- Used for data analysis and finding patterns
- Always chooses components that maximize variance
- Compress your data while retaining most of the information

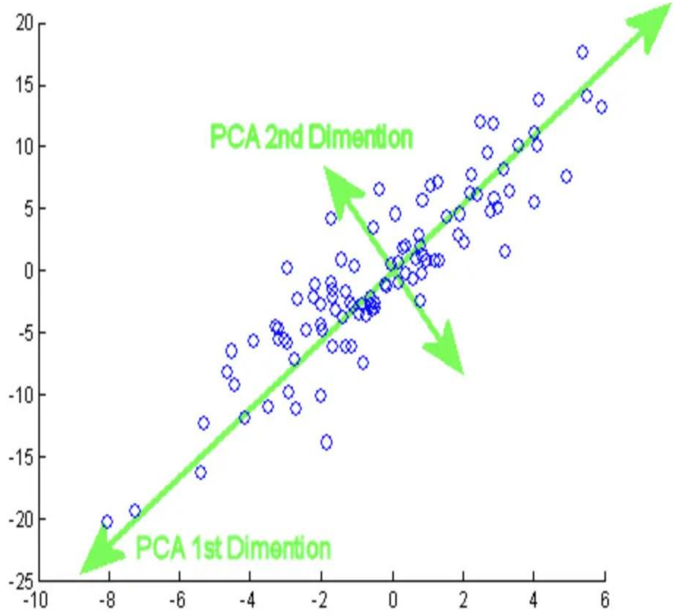


Fig.7 PCA Dimensionality Reduction

http://bit.ly/3G51zqM

Key Differences

	PCA	LDA
Objective	Maximize Variance	Maximize Separation
Type	Unsupervised	Supervised
Outputs	Principal components	Discriminant Functions
Use case	EDA	Classification

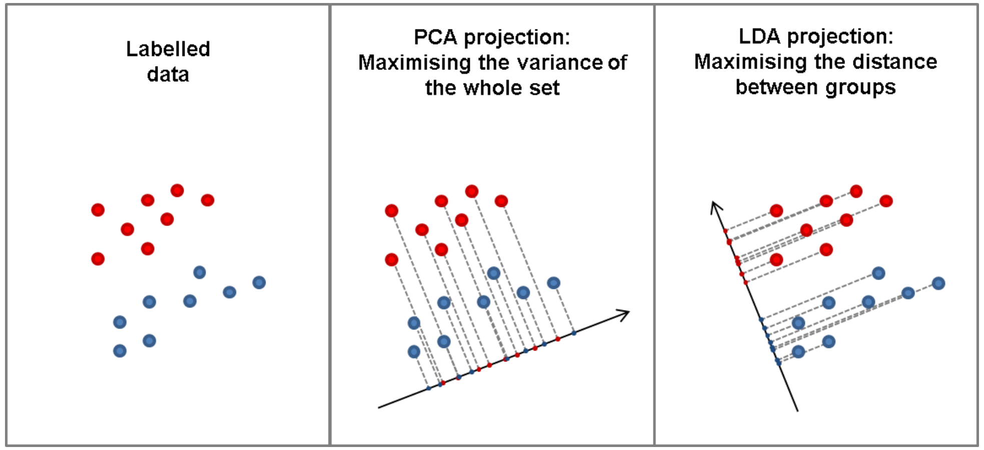


Fig.8 PCA Vs LDA
https://bit.ly/3G4geTh

Assumptions

- **Normality** - LDA assumes that the features within each class are normally distributed [2].
- **Homoscedasticity** – Each feature should have the same amount of variance.
- **Independence of Features** – There should be little or no multicollinearity.
- **Linearity** – Relation between features and class labels should be linear.

LDA Implementation

- Fisher's Discriminant Ratio = $\frac{(\mu_1 - \mu_2)^2}{{s_1}^2 + {s_2}^2}$ [4]

- **Step 1** – Compute 'Within Class Scatter (S_w)' and 'Between Class Scatter (S_B)'
  - $S_W = \sum_{i=1}^{c} s_i$ where $S_i = \sum_{x \in C_i}^{n} (x - \mu_i)(x - \mu_i)^T$
  - $S_B = \sum_{i=1}^{c} N_c(\mu_i - \mu)(\mu_i - \mu)^T$

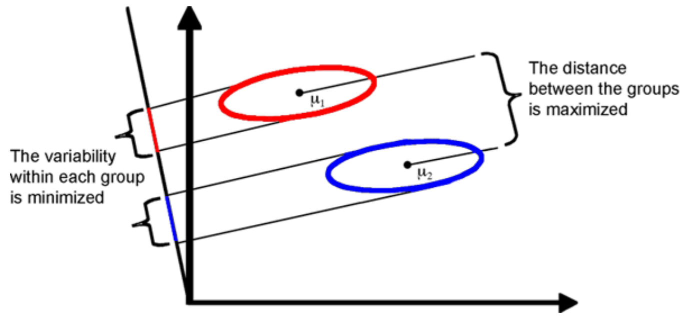


Fig.9 LDA Projection Vector

https://bit.ly/47GkNPz

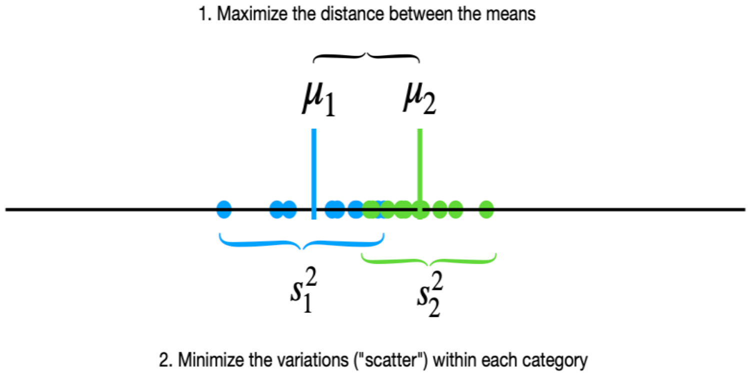


Fig.10 Points projected on vector

https://bit.ly/3QK7aHV

University of Windsor

LDA Implementation

- **Step 2** – Compute the Eigen Values and Eigen Vectors for
  - $S_W{}^{-1}.S_B\vec{v} = \lambda\vec{v}$
  - $\left|S_W{}^{-1}.S_B - \lambda I\right| = 0$
- **Step 3** – Select the top Eigen Vectors and form the Transformational Matrix
- **Step 4** – Transform the Input Data as
  - $Y = W^T.\mathrm{X}$

$\mu_1\mu_2$	Mean
s_1s_2	Scatter
N_c	Number of classes
λ	Eigen Values
W	Transformational matrix
Y	Transformed Output
X	Transformed Input
$\vec{v}$	Eigen Vector

Table2. Variables and their Meanings.

Program Implementation

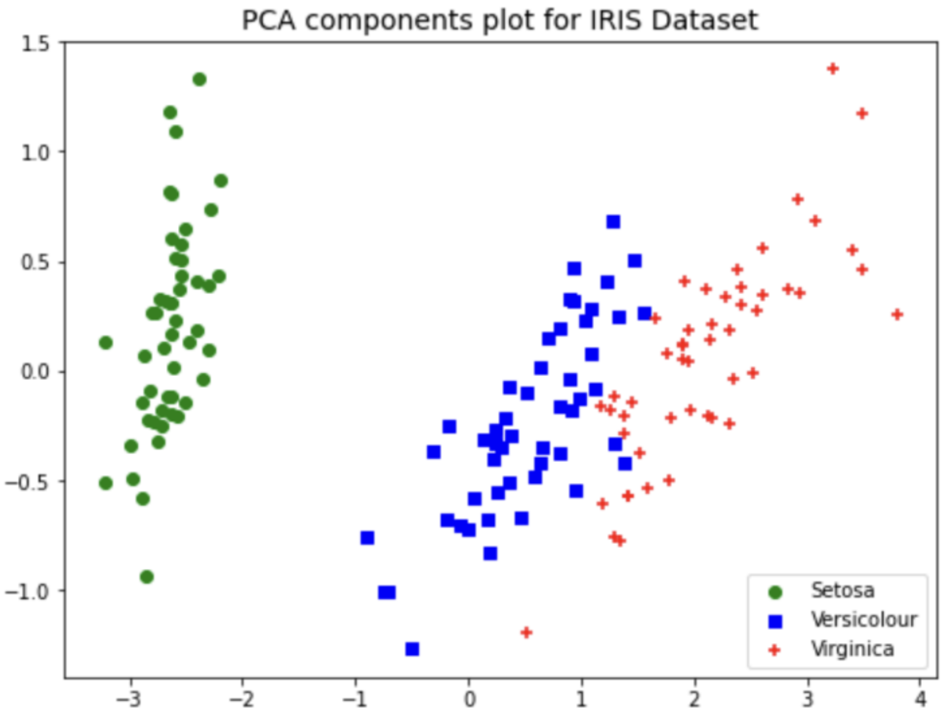


Fig.11 Iris dataset plot on PCA Component Axes

https://bit.ly/47yFoVN

```
import pandas as pd
import numpy as np
from sklearn.decomposition import PCA
import matplotlib.pyplot as plt

from sklearn import datasets
#
# Load IRIS dataset
#
iris = datasets.load_iris()
#
# Create a dataframe from IRIS dataset
#
df = pd.DataFrame(iris.data, columns=["sepal_length", "sepal_width",
"petal_length", "petal_width"])
df["class"] = iris.target
#
# Create PCA transformed dataset with dimensionality
# reduced to 2; n_components = 2
#
pca = PCA(n_components=2)
X_pca = pca.fit(df.iloc[:, 0:4]).transform(df.iloc[:, 0:4])
#
# Create plot from transformed dataset
#
plt.figure(figsize=(8,6))

plt.scatter(X_pca[0:50,0], X_pca[0:50,1], color='green', marker='o',
label='Setosa')
plt.scatter(X_pca[50:100,0], X_pca[50:100,1], color='blue', marker='s',
label='Versicolour')
plt.scatter(X_pca[100:150,0], X_pca[100:150,1], color='red', marker='+',
label='Virginica')

plt.title("PCA components plot for IRIS Dataset", fontsize=14)
plt.legend()
plt.show()
```

Program Implementation

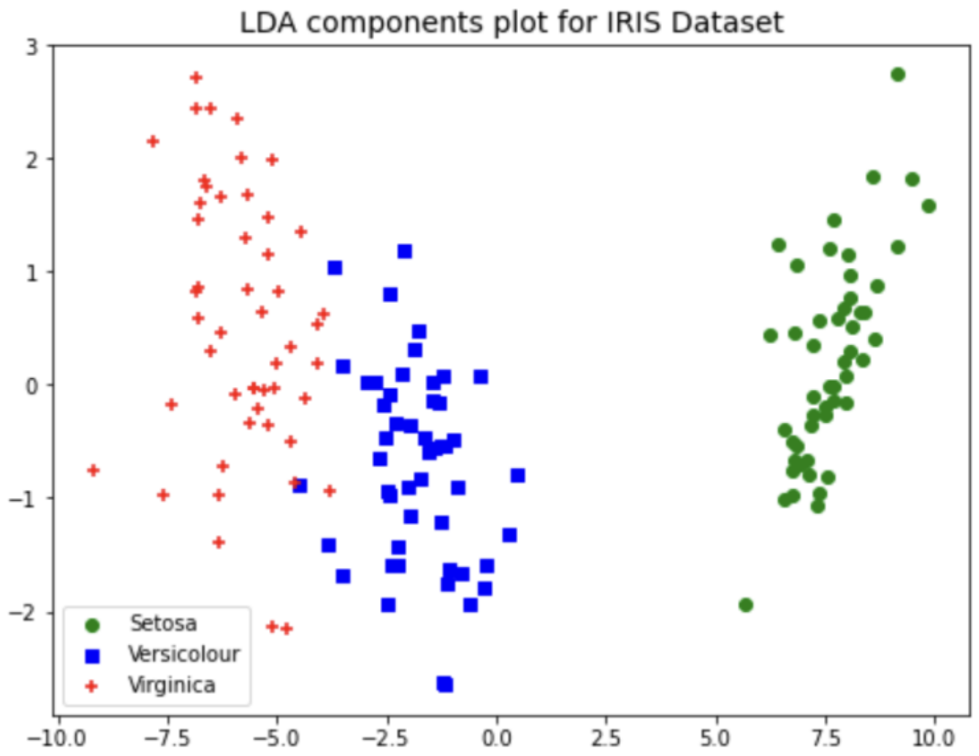


Fig. 12 Iris dataset plot on LDA Component Axes

https://bit.ly/47yFoVN

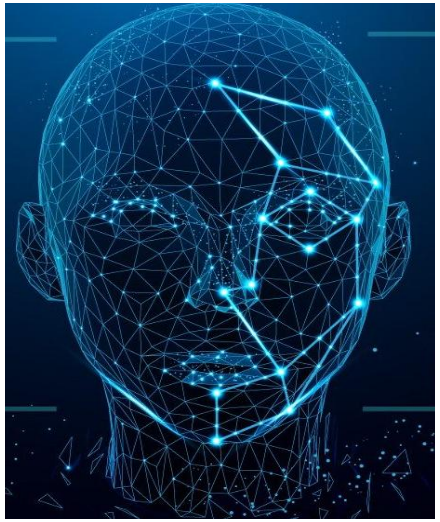
```python
import pandas as pd
import numpy as np
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
import matplotlib.pyplot as plt

from sklearn import datasets
#
# Load IRIS dataset
#
iris = datasets.load_iris()
#
# Create a dataframe from IRIS dataset
#
df = pd.DataFrame(iris.data, columns=["sepal_length", "sepal_width",
"petal_length", "petal_width"])
df["class"] = iris.target
#
# Create LDA transformed dataset with dimensionality
# reduced to 2; n_components = 2
#
lda = LinearDiscriminantAnalysis(n_components=2)
X_lda = lda.fit(df.iloc[:, 0:4], df.iloc[:, -1]).transform(df.iloc[:,
0:4])
#
# Create plot from transformed dataset
#
plt.figure(figsize=(8,6))

plt.scatter(X_lda[0:50,0], X_lda[0:50,1], color='green', marker='o',
label='Setosa')
plt.scatter(X_lda[50:100,0], X_lda[50:100,1], color='blue', marker='s',
label='Versicolour')
plt.scatter(X_lda[100:150,0], X_lda[100:150,1], color='red', marker='+',
label='Virginica')

plt.title("LDA components plot for IRIS Dataset", fontsize=14)
plt.legend()
plt.show()
```


University of Windsor

Applications of LDA

- To classify the patient disease(Either increase or decrease pace of treatment)
- Drug suitability
- Early detection of disease(EHR data)

- Face recognition & Biometrics
- Each face is represented as the combination of a number of pixel values.
- LDA is used to minimize the number of features



Marketing and Customer Segmentation:

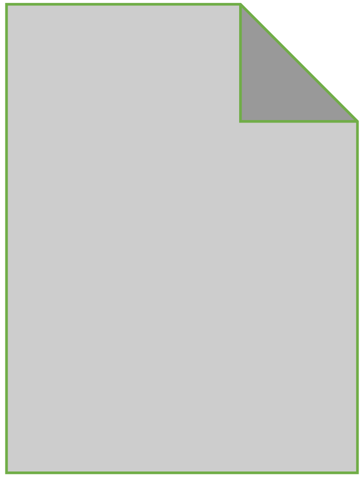
- By classifying customers into different segments based on their purchasing behavior, demographic information, and preferences.

Applications of LDA

Remote Sensing and Image Analysis:

- LDA can be used for classifying land cover types in satellite images or aerial photographs, which helps to differentiate between different types of terrain, vegetation, or land use.

Applications of LDA

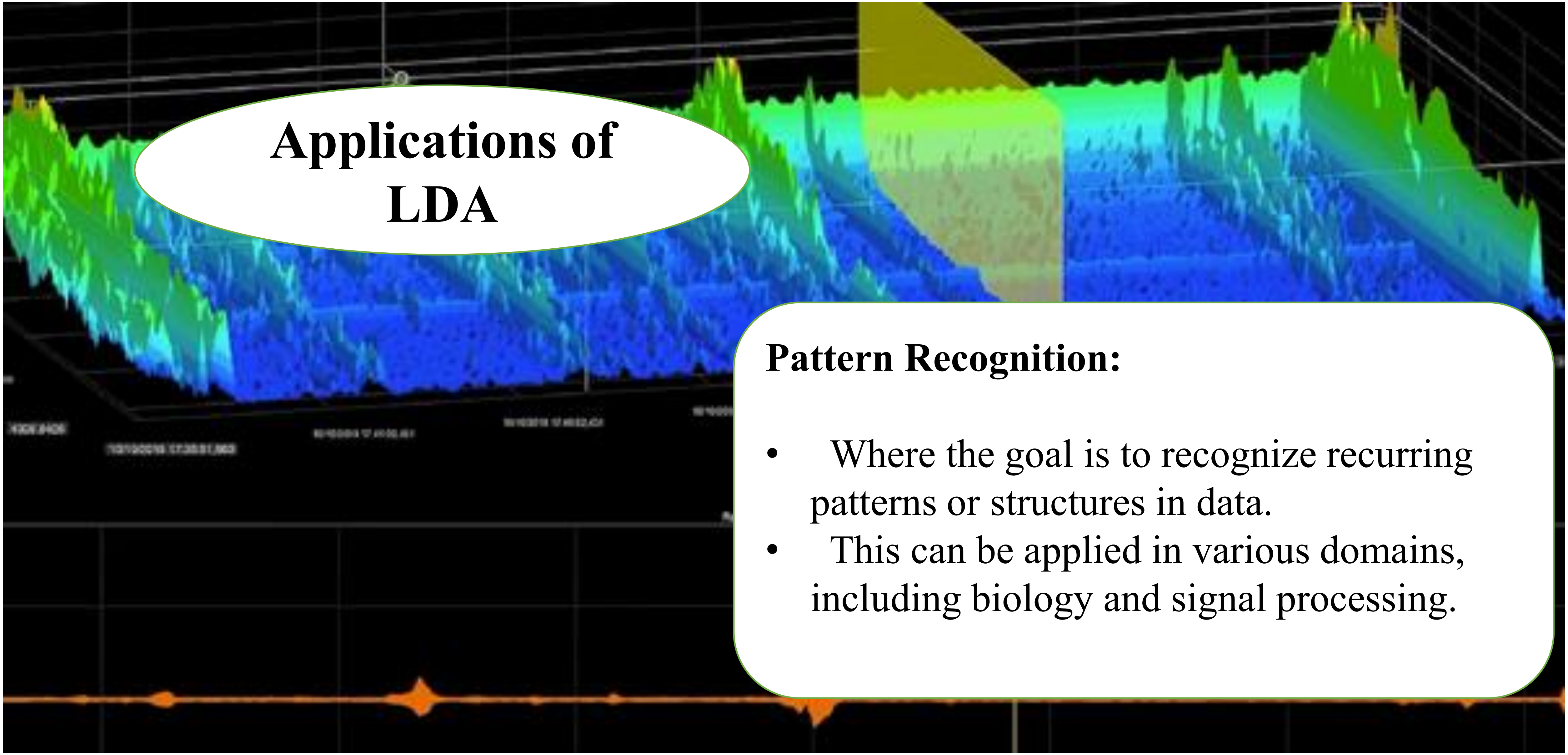


Quality Control and Manufacturing:

- Can assist in identifying defects in products by classifying items as defective or non-defective.
- Useful in industries like manufacturing and production.

Document Classification:

- Can categorize documents into different classes or topics.
- For instance, spam and non-spam categories of E-mails or news articles into different sections.



Applications of LDA

Pattern Recognition:

- Where the goal is to recognize recurring patterns or structures in data.
- This can be applied in various domains, including biology and signal processing.

University of Windsor

Advantages and limitations of LDA

Advantages

- Dimensionality Reduction with Class Separation-Simple and computationally efficient algorithm.
- Utilizes Class Information
- Works Well for the number of features >> the number of training samples.
- Data Visualization
- Robust to Outliers due to its reliance on class means and variances rather than individual data points.

Limitations

- Sensitive to Class Distribution(Gaussian distribution).
- Prone to Overfitting(Curse of dimensionality).
- Doesn't Handle Nonlinear Relationships.
- Requires Well-Defined Classes(supervised technique and relies on class labels for training).
- Doesn't Incorporate Feature Interaction.
- May Not Capture Complex Patterns(Non liner or neural networks).

Extensions to LDA

Non-linear Discriminant analysis

- **Quadratic Discriminant Analysis (QDA):** Each class uses its own estimate of variance (or covariance when there are multiple input variables).

- **Flexible Discriminant Analysis (FDA):** Where non-linear combinations of inputs are used such as splines.

- **Regularized Discriminant Analysis (RDA):** Introduces regularization into the estimate of the variance (actually covariance), moderating the influence of different variables on LDA.

Conclusion

-LDA
supervised learning
Aims to find LD to represent axes that maximize separation between different classes of data

-PCA
unsupervised learning
-Aims to find PC to maximize the variance

-Both are Dimensionality Reduction techniques

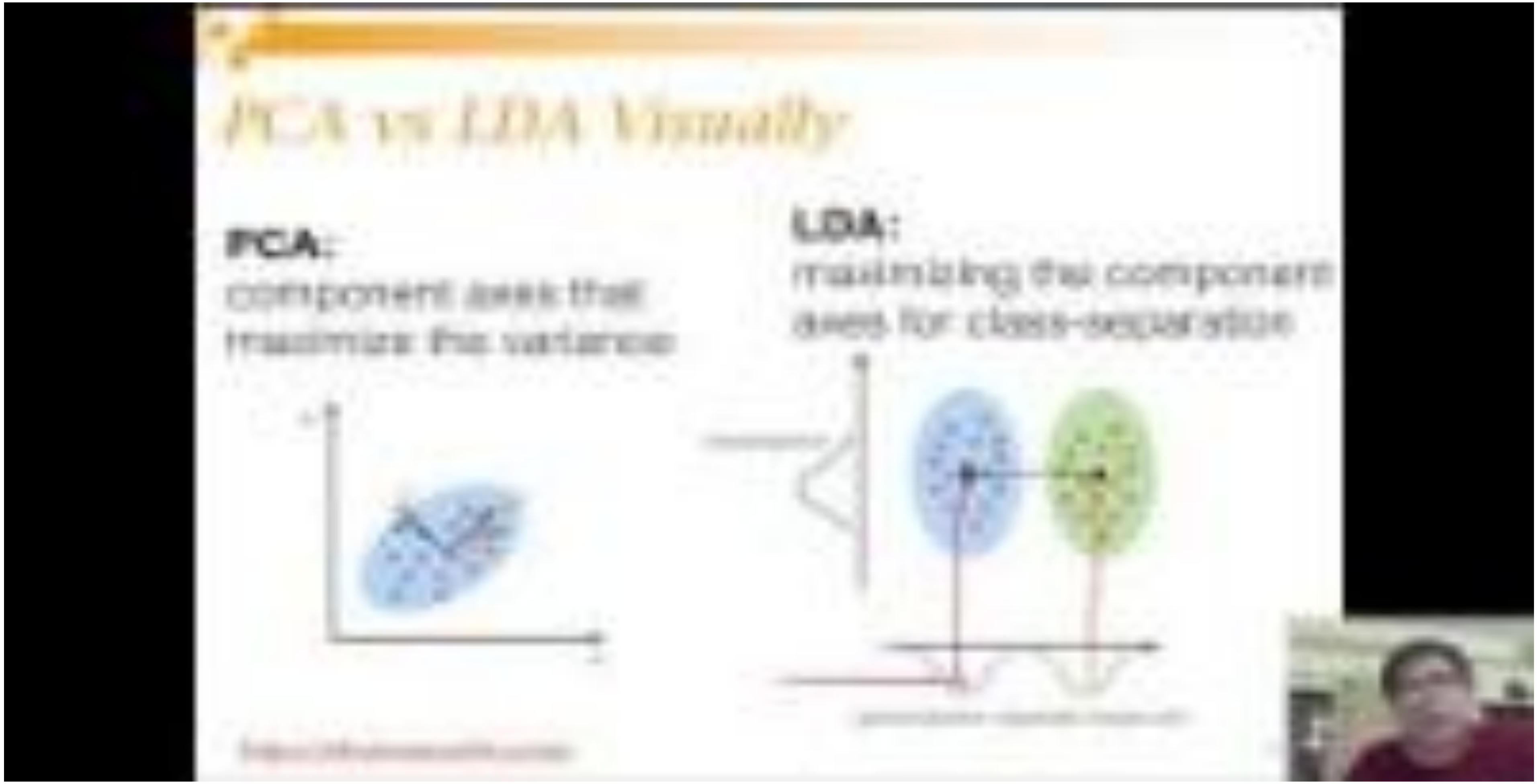
-LDA fails to create new axes which seperates 2 class linearly, in cases where mean of distribution is shared

LDA is much more suitable for multi-class classification tasks compared to PCA.

References

[1] “Principal component analysis,” Wikipedia, https://en.wikipedia.org/wiki/Principal_component_analysis (accessed Nov. 17, 2023).
[2] “PCA vs LDA Differences, plots, examples,” Analytics Yogi, https://vitalflux.com/pca-vs-lda-differences-plots-examples/ (accessed Nov. 17, 2023).
[3] Seldon, “Supervised vs unsupervised learning explained,” Seldon, https://www.seldon.io/supervised-vs-unsupervised-learning-explained#:~:text=Supervised%20machine%20learning%20is%20generally,the%20need%20for%20labelled%20data. (accessed Nov. 17, 2023).
[4] Scientific Computing and Imaging Institute, https://sci.utah.edu/~shireen/pdfs/tutorials/Elhabian_LDA09.pdf (accessed Nov. 17, 2023).

For Better Understanding



https://www.youtube.com/watch?v=D2HArUvOQaw&ab_channel=SaptarsiGoswami

Thank you!