

ADVANCED STATISTICS (Week 3)

DSBA CURRICULUM DESIGN

FOUNDATIONS

**Python For Data
Science**

**Statistical Methods
for Decision
Making**

CORE COURSES

**Advanced
Statistics(Week-3/5)**

Data Mining

**Predictive
Modelling**

Machine Learning

**Time Series
Forecasting**

Data Visualization

SQL

DOMAIN APPLICATIONS

**Financial Risk
Analytics**

**Marketing Retail
Analytics**

LEARNING OBJECTIVE OF THIS MODULE

- ANOVA
- EDA
- PCA

LEARNING OBJECTIVES OF THIS SESSION -

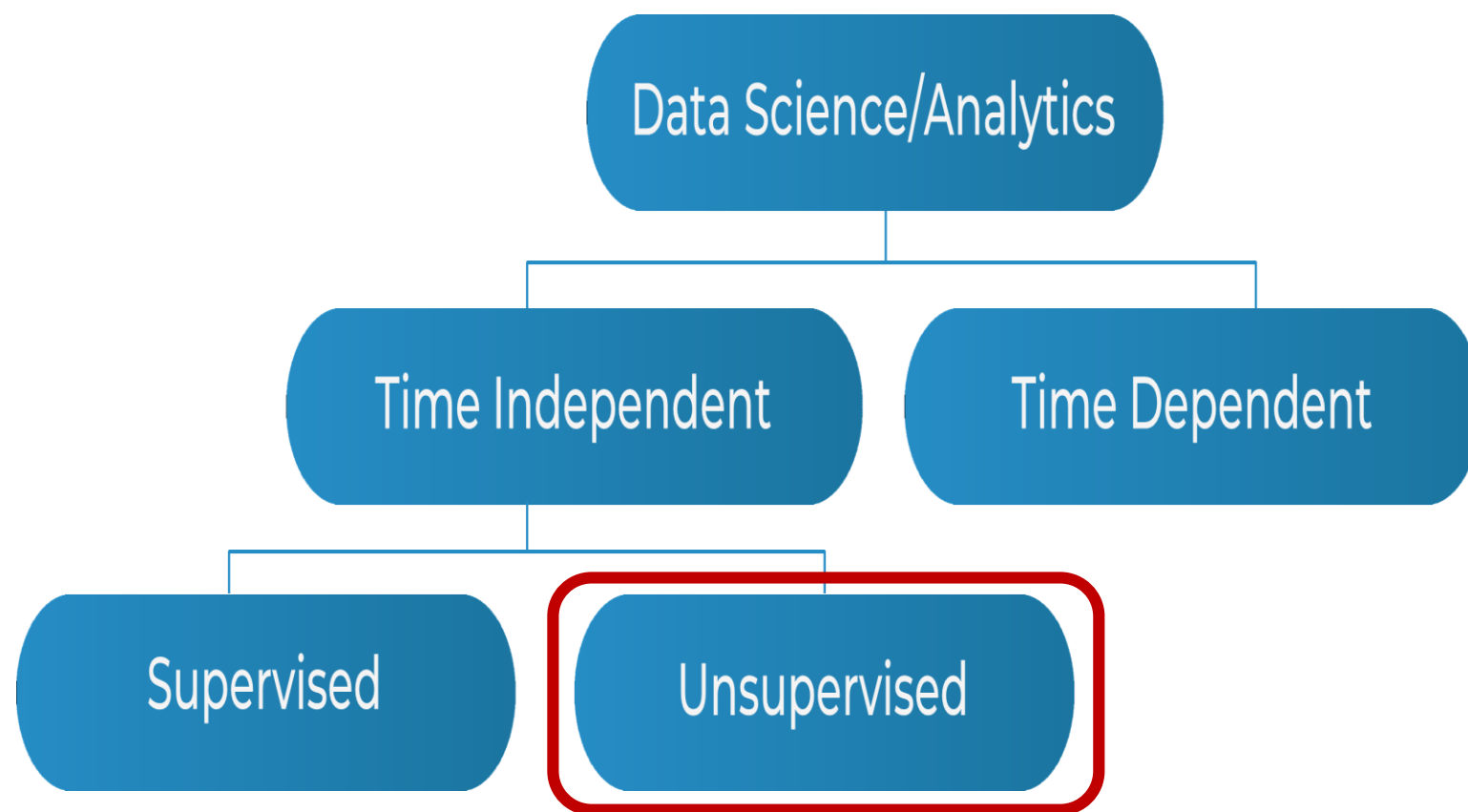
- Covariance Matrix
- Eigen Value
- Variance and Cumulative Variance
- Scree Plot

TRY ANSWERING THE FOLLOWING

- Can we use PCA for dimensionality reduction?
- What is the purpose of using Scree Plot in PCA?
- Does different Principal Components orthogonal to each other?

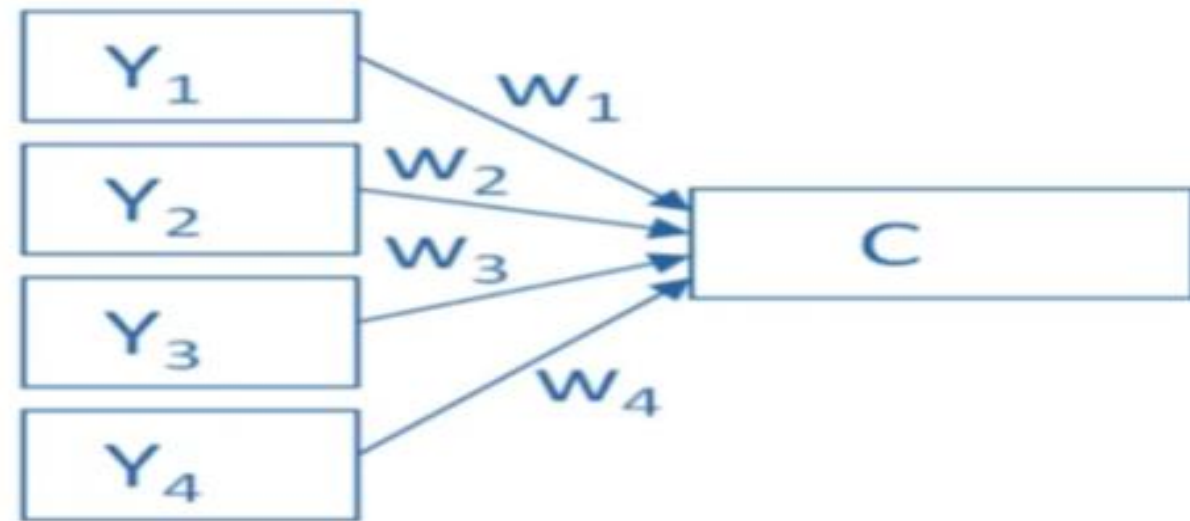


BROAD OVERVIEW: Principal Component Analysis (PCA)



- Dimension reduction technique
- Pattern Recognition (Based on Features)
- Resolve Multicollinearity

PCA Technique



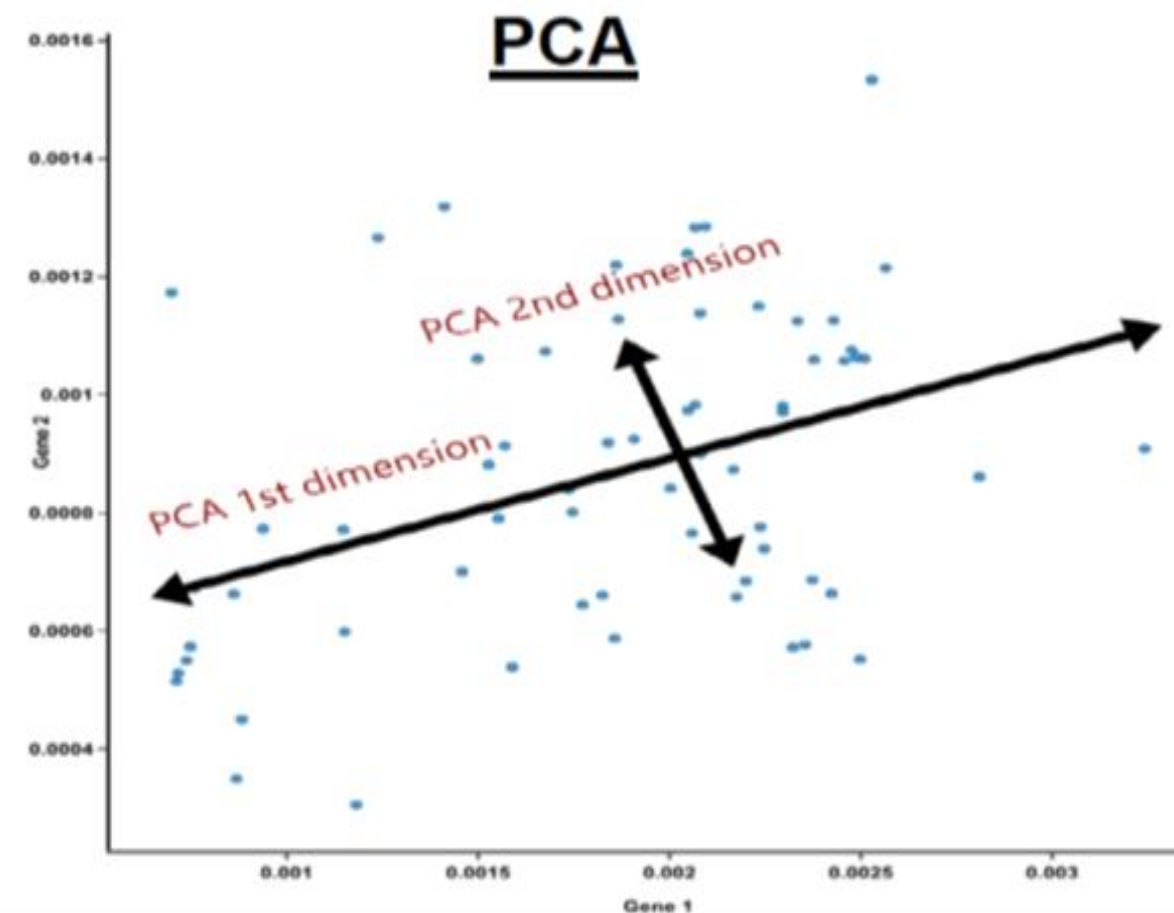
- *Linear combination of variables*

$$C = w_1(Y_1) + w_2(Y_2) + w_3(Y_3) + w_4(Y_4)$$

C: Component

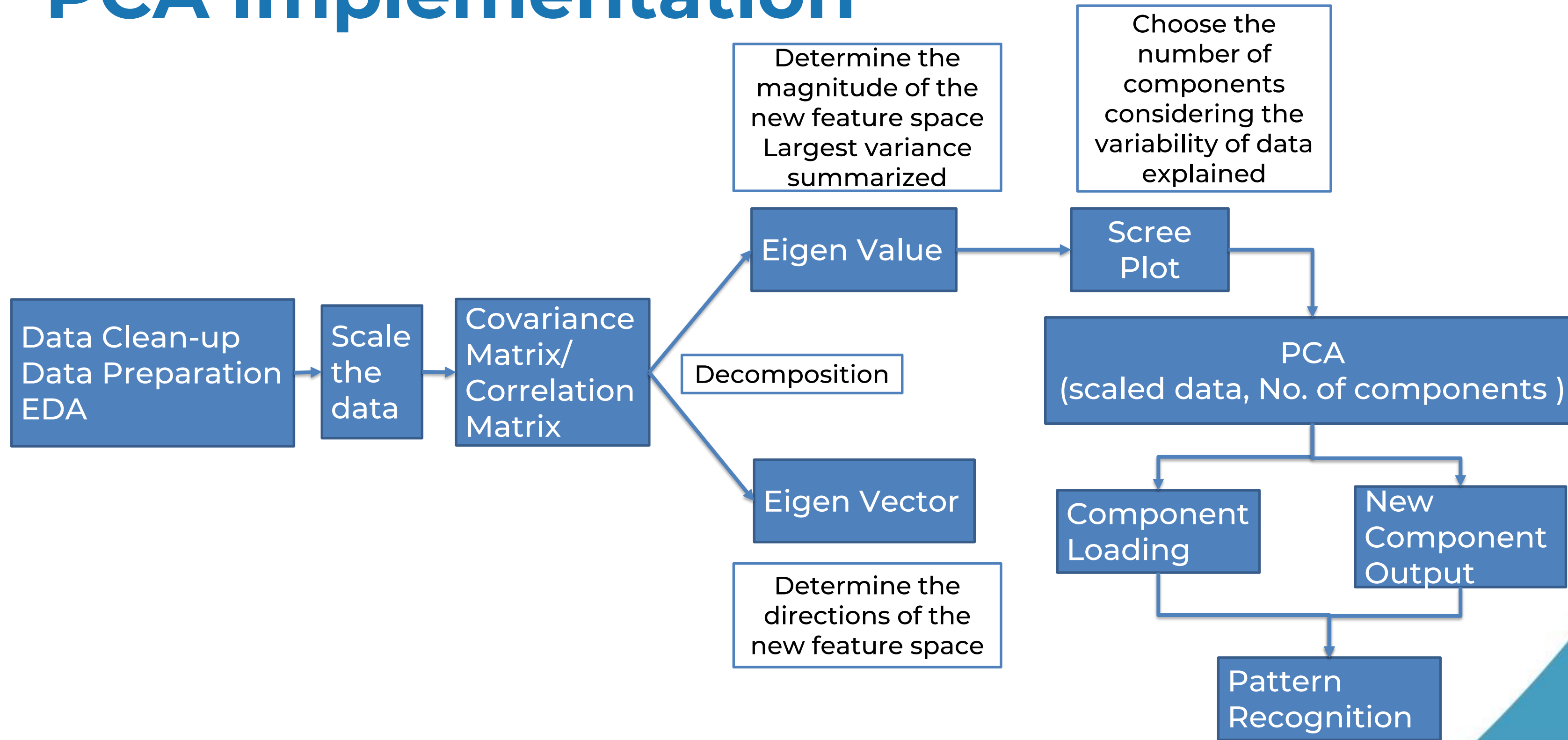
w_1, w_2, w_3, w_4 : pca components loadings

Y_1, Y_2, Y_3, Y_4 : Features



- PCA1 *accounts for as much variation* in the data as possible
- PCA2 accounts for as much as remaining variation in the data as possible, with the constraint that the *correlation between PCA1 and PCA2 is zero*
- All subsequent components have the same property

PCA Implementation



Let's Learn Together – A Unique Platform for Peer to Peer Learning

Next Week's Theme:

Mind-Map of Advance Statistics Course

Reference Link-<https://www.mindmapping.com/>



Benefits of Creating Mind-Map:

- ❖ Quick Revision of the course
- ❖ This document will prove very handy later in the course
- ❖ Opportunity to show your creativity

What all can be discussed in a Discussion forum?

- ❖ Analytical Concepts
- ❖ Issues in Code
- ❖ Real Time/Industry Examples

Case Study: PCA on Loan Transactions



Lending is one of the critical functions of any financial institution. Customers are provided loans across different products at competitive interest rates for an acceptable tenure. There is always a risk that a customer may

- default on the loan or
- may try and repay the loan in advance

which leads to financial losses to the business.

There may not be enough data or evidence available from the past which will help the firm to predict and mitigate the above possible risk.

However, it does have information about the customer demographics, loan details, EMI transactions etc. which may contribute to 30+ features related to a customer's loan account.

Let's apply PCA on this high dimension data, identify patterns and help the bank with some useful information to monitor the risk



ANY QUESTIONS



HAPPY LEARNING