greatlearning
Learning for Life

ADVANCE STATISTICS (Week 3)

greatlearning
Learning for Life

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

greatlearning
Learning for Life

LEARNING OBJECTIVE OF THIS MODULE

- ANOVA
- EDA
- PCA

greatlearning
Learning for Life

LEARNING OBJECTIVES OF THIS SESSION -

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

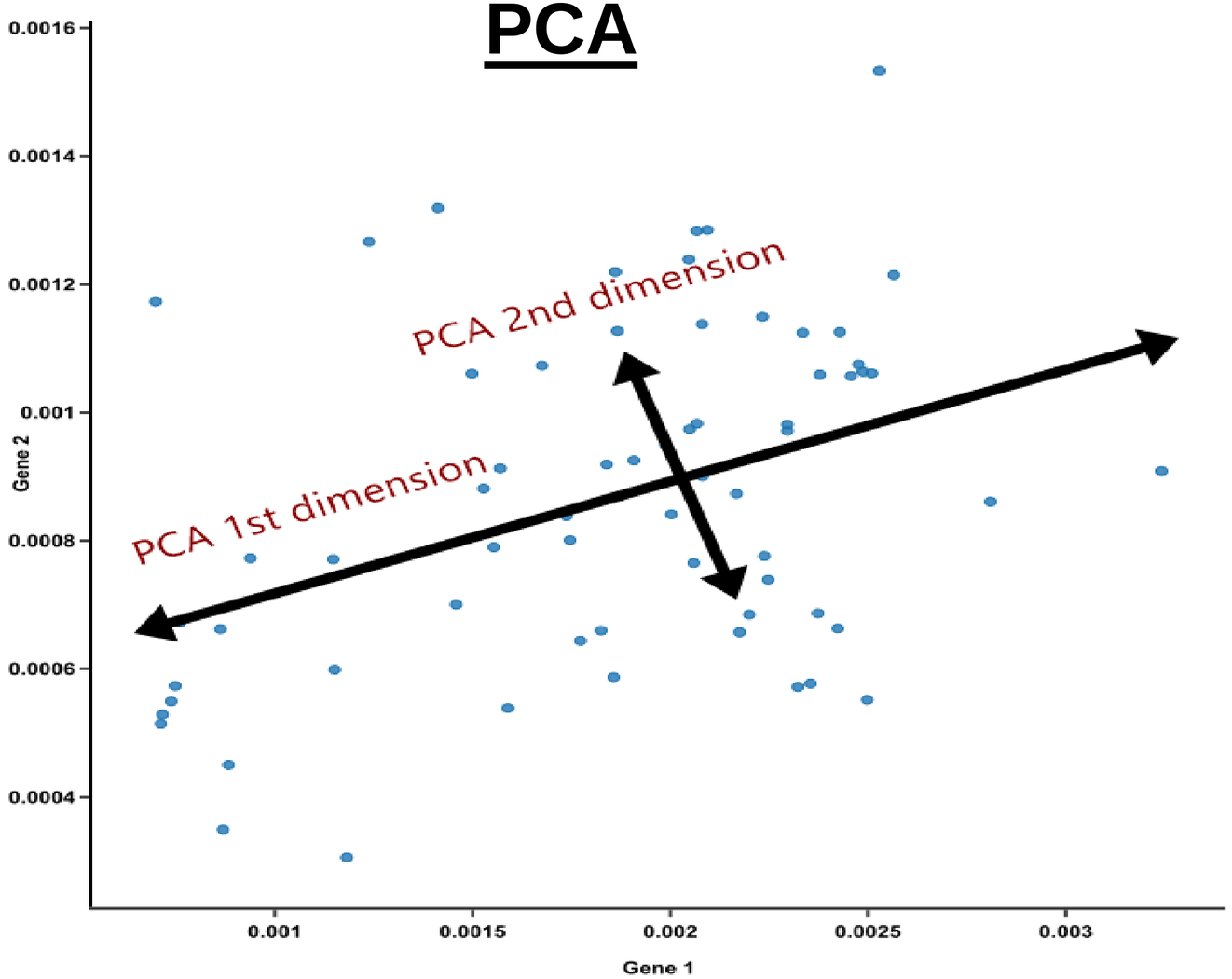
greatlearning
Learning for Life

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?

greatlearning
Learning for Life

BROAD OVERVIEW



PCA to devise IPL's bidding strategy

Real game begins much before it begins on the pitch.

Indian Premier League (IPL) is a professional Twenty20 championship cricket league in India that has become very popular among cricket fans worldwide. There are several ways that a franchise acquires players. One of these methods is to buy players at player auction. Obviously, franchises pay higher salaries for quality players based on their performance.

Use of PCA-When considering batsmen, the goal in limited-overs cricket, like Twenty20, is to score as many as runs as possible using as few balls as possible. Therefore, variables considered are Runs, Batting Average, Strike Rate, Number of Fours, Number of Sixes, Number of Centuries/Half Centuries etc. But the question is how do we decide which variable is more important and finally can we generate a score that helps us decide which player to bet on?

The answer can be achieved using PCA. Find a linear combination of these variables and use each player's performance to generate a score.

Reference-http://jse.amstat.org/v21n3/scariano.pdf

greatlearning
Learning for Life

PCA for Designing Survey Questionnaire

Are you asking the right questions?

What single measure describes your overall health?
For example, take questions pertaining to height, weight, smoking habit, exercise routine, blood pressure, diabetes, heart disease etc in a healthcare survey. Individually, the responses to these questions are too specific to say much about your overall health. But taken together, they can provide a more comprehensive measure of your wellbeing, which is the common factor that researchers are truly interested in.

Instead, many researchers will ask a series of health-related questions and perform a factor analysis, which generates a standardized score of health. It helps us find out whether variables (or in the case of surveys, questions) are correlated with one another or with some other variable or concept. Statisticians call these related variables common factors.
We can determine the relationships between groups of variables by lumping together the ones that are that are strongly correlated, making them into common factors. This is the basis of factor analysis, which is often used in the fields of psychology, health, and political science.
https://www.surveygizmo.com/resources/blog/factor-analysis/

greatlearning
Learning for Life

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

greatlearning
Learning for Life

Case Study- (Silhouette of Vehicles)

The data contains features extracted from the silhouette of vehicles in different angles. Four "Corgie" model vehicles were used for the experiment: a double decker bus, Cheverolet van, Saab 9000 and an Opel Manta 400 cars. This particular combination of vehicles was chosen with the expectation that the bus, van and either one of the cars would be readily distinguishable, but it would be more difficult to distinguish between the cars.

The purpose is to classify a given silhouette as different vehicle, using a set of features extracted from the silhouette.

Attribute Information:
All the features are geometric features extracted from the silhouette.
All are numeric in nature.

greatlearning
Learning for Life

ANY QUESTIONS

greatlearning
Learning for Life

HAPPY LEARNING