Principal Component Analysis with python #2262

PRESENTED BY: DEEPTHI M

BATCH NUMBER: 05

SERIAL NUMBER: 198

Dimension Reduction:

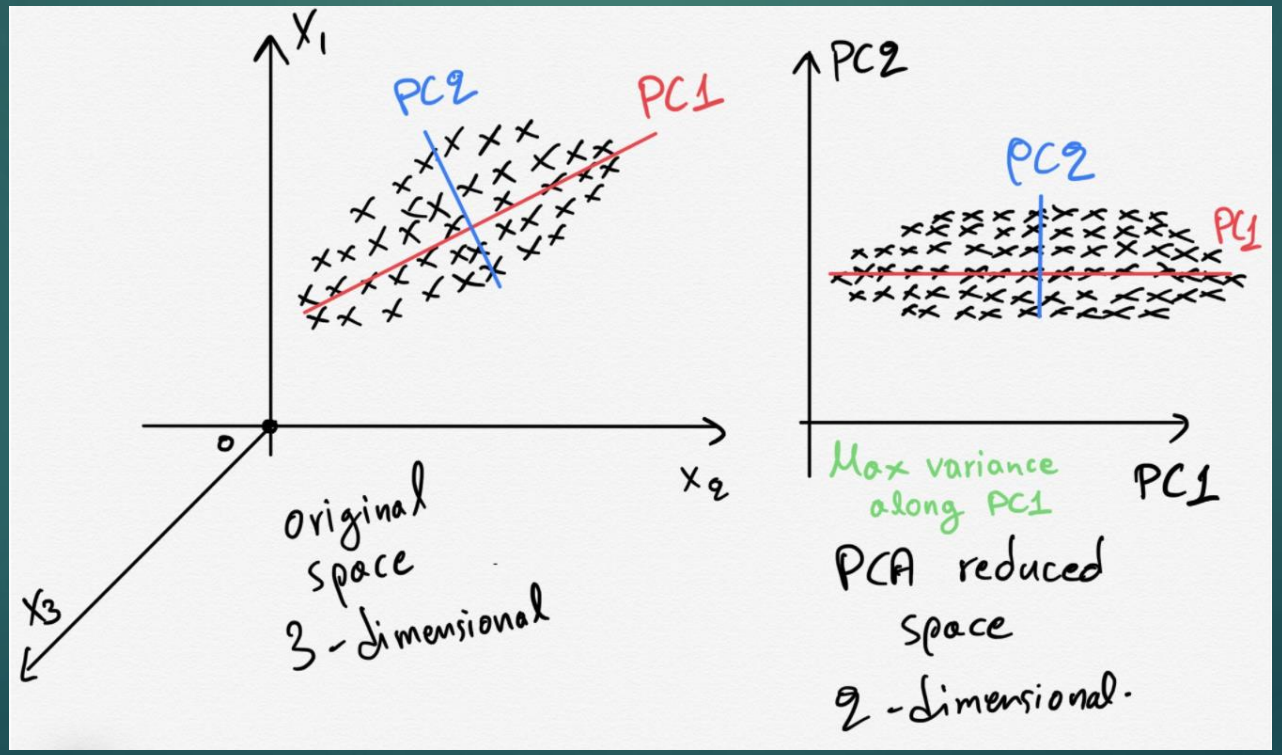
- Principal Component Analysis (PCA)
- Singular Value Decomposition (SVD)

x1	x2	x3	x4	x5	x6	x7	x8

x1	x2	x3

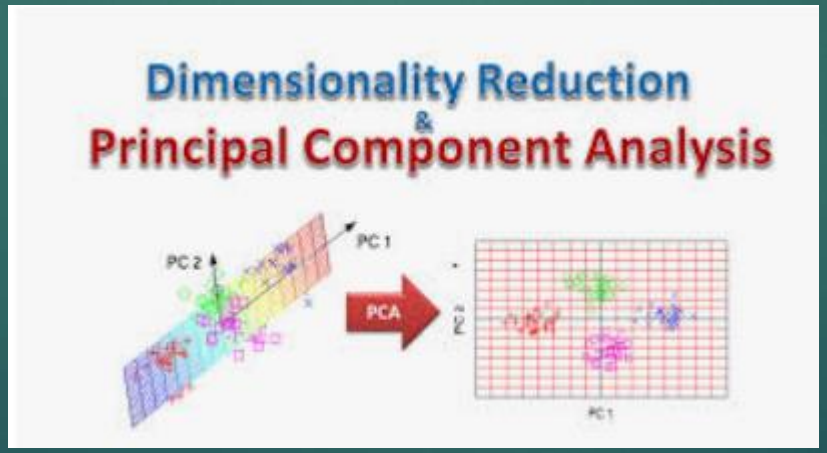
Principal Component Analysis (PCA)

- **PCA** is a statistical method that is used to convert a set of correlated variables to a set of uncorrelated variables.
- It is a unsupervised machine learning model.
- Broadly, used as a EDA(Exploratory Data Analysis) method.



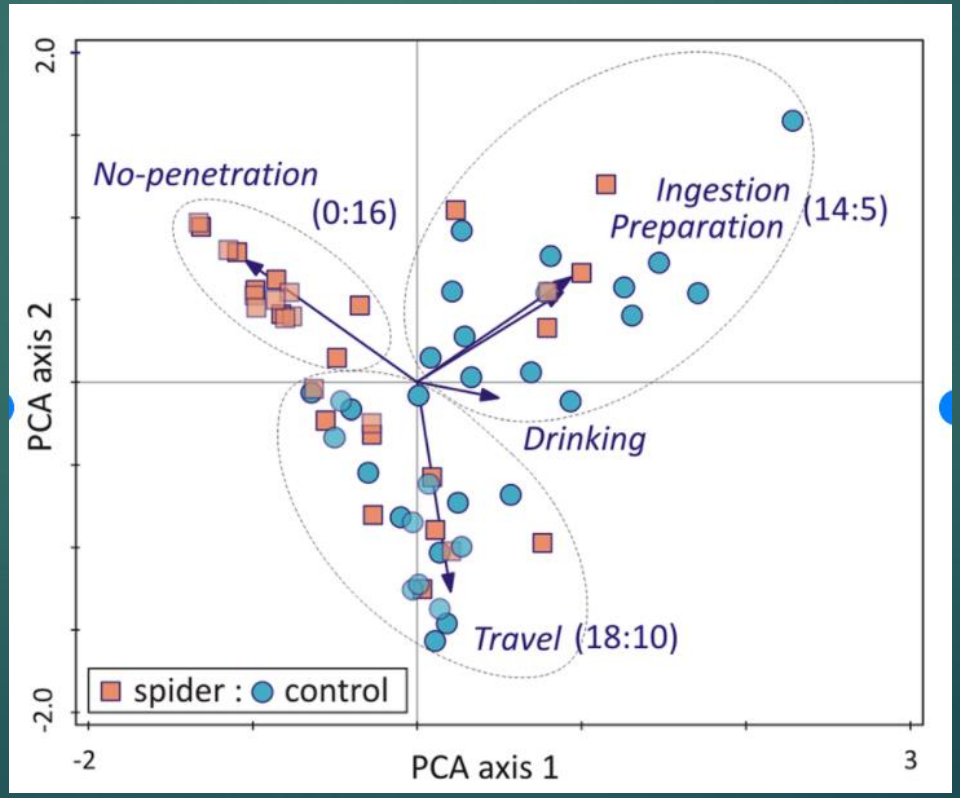
Used to convert large number of features into a fewer features with some loss of information:

- Identify the relationship between columns
- Visualize the multivariate data using 2 Principal Components
- It will help condense the features



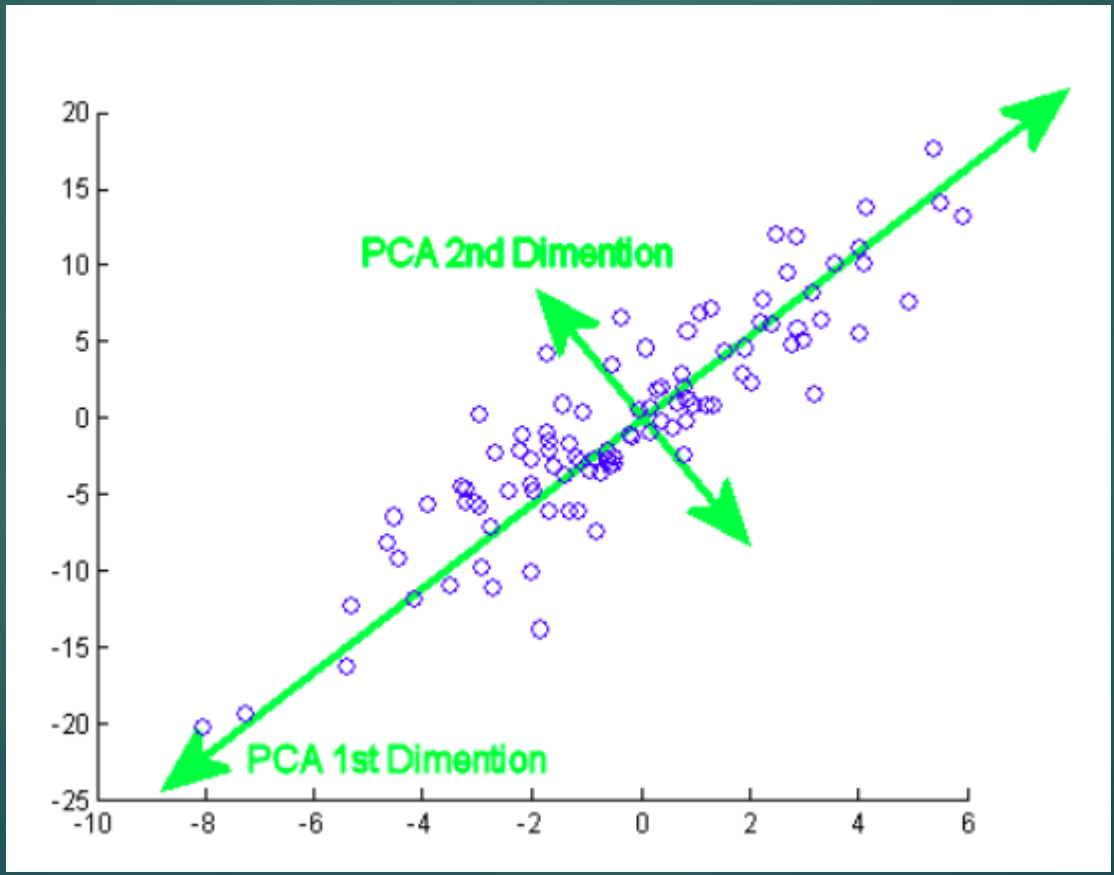
When/Why to use PCA:

- PCA technique is particularly useful in processing data where **multi-colinearity** exists between the **features/variables**.
- PCA can be used when **the dimensions of the input features are high** (e.g. a lot of variables).
- PCA can be also used for **denoising** and **data compression**.



Some of the Applications of PCA:

- Computation is fast
- Image compression



Dimensionality reduction

- Pros
  - reflects our intuitions about the data
  - allows estimating probabilities in high-dimensional data
    - no need to assume independence etc.
  - dramatic reduction in size of data
    - faster processing (as long as reduction is fast), smaller storage
- Cons
  - too expensive for many applications (Twitter, web)
  - disastrous for tasks with fine-grained classes
  - understand assumptions behind the methods (linearity etc.)
    - there may be better ways to deal with sparseness

I hope now we have an overview on the topics:

- PCA

Thank you I hope you have enjoyed!!!