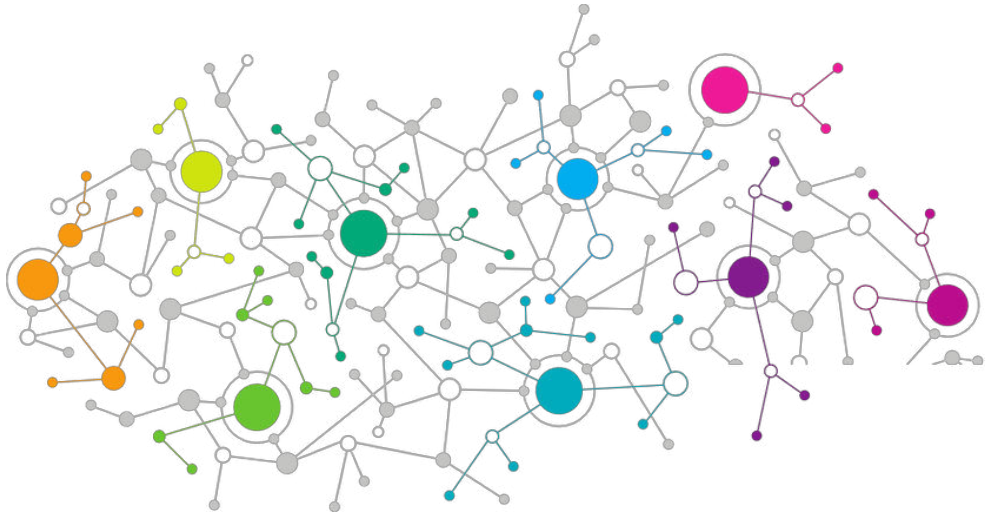
IntelliPaat

Data Science

Module-9.1

Principal Component Analysis



Agenda

01 DIMENSIONALITY REDUCTION

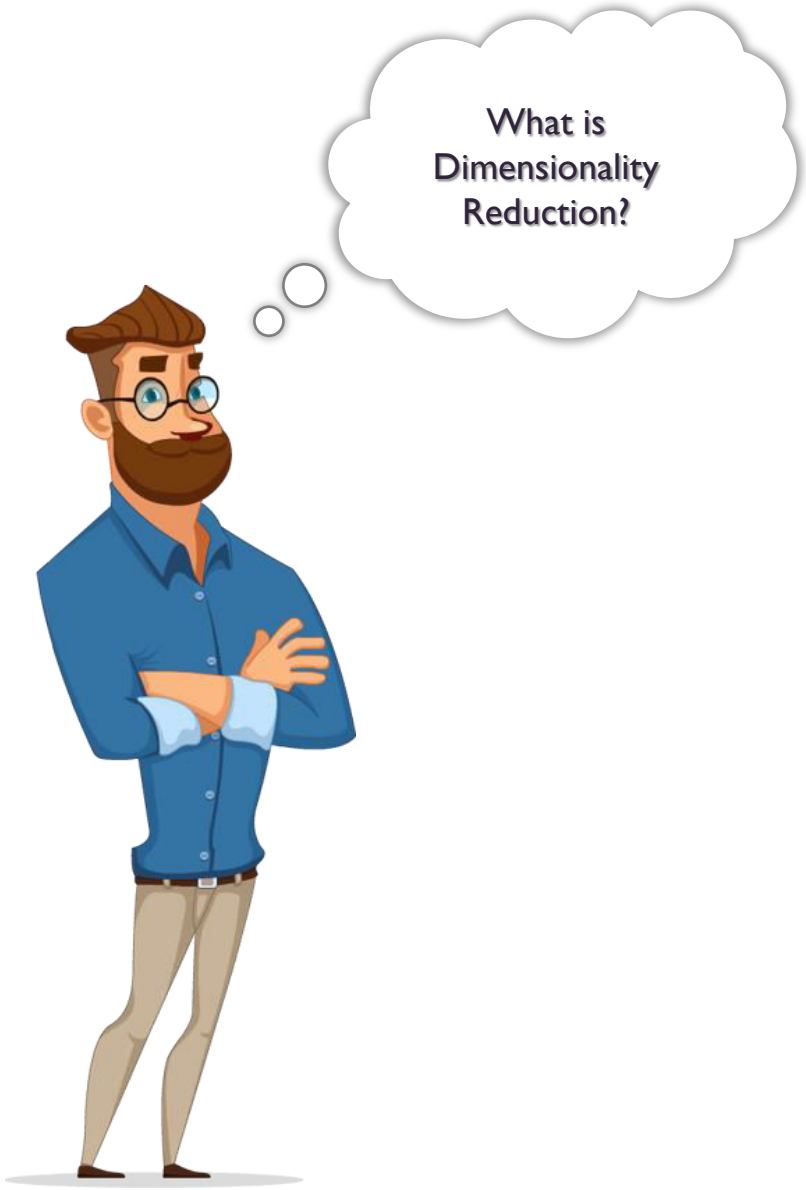
02 TYPES OF DIMENSIONALITY REDUCTION

03 PRINCIPAL COMPONENT ANALYSIS

04 WORKING MECHANISM OF PCA

Dimensionality Reduction

Dimensionality Reduction

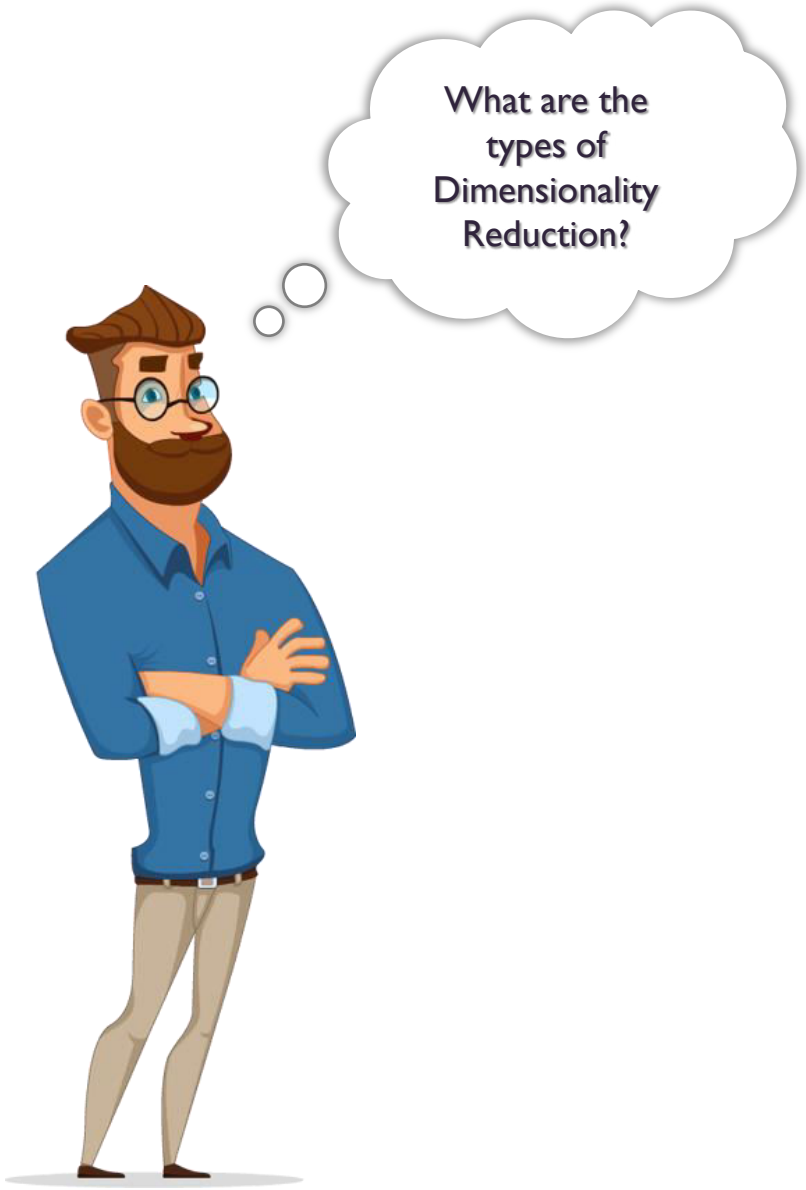


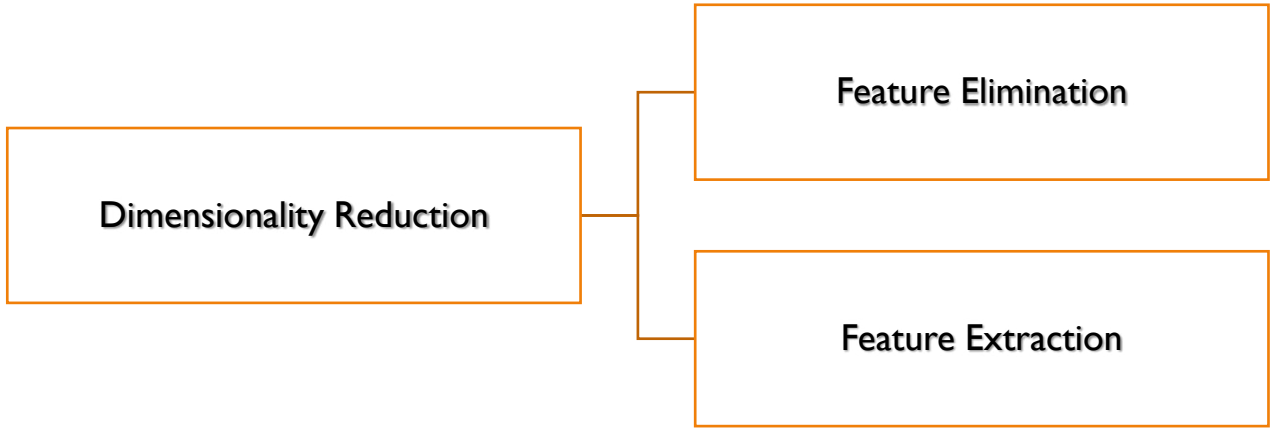
Converting data set of vast dimensions into data with lesser dimensions

Reduce the complexity of data by keeping the relevant structure

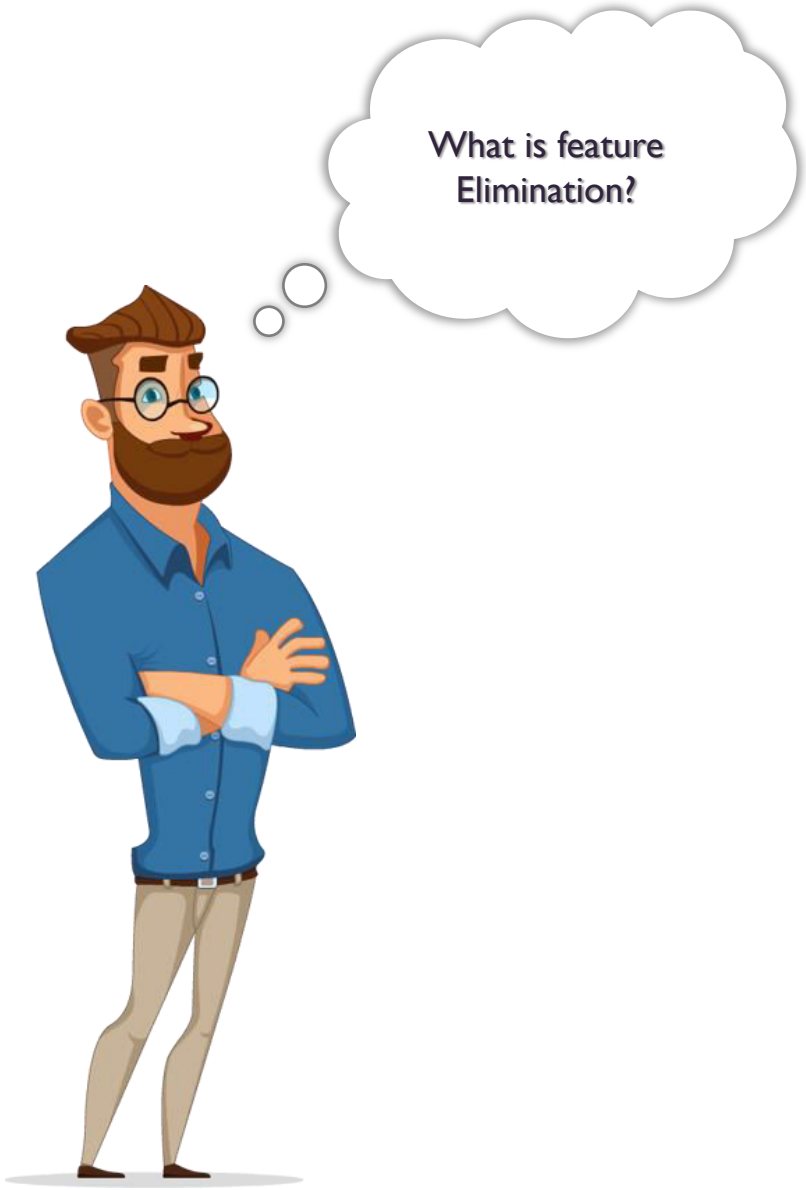
Types of Dimensionality Reduction

Dimensionality Reduction





Dimensionality Reduction



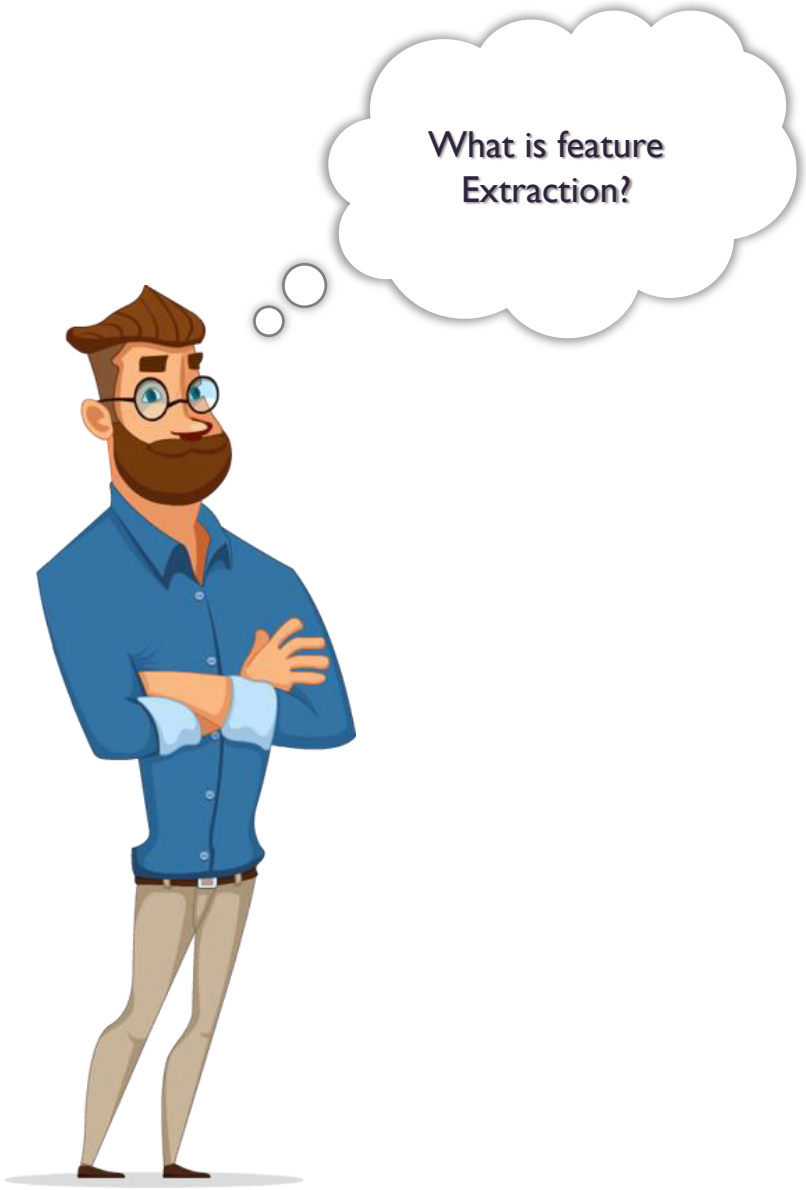
Removing some variables completely if,

- Redundant with other variable
- Not providing any new information

Sets smaller dataset

Might lose some data

Dimensionality Reduction

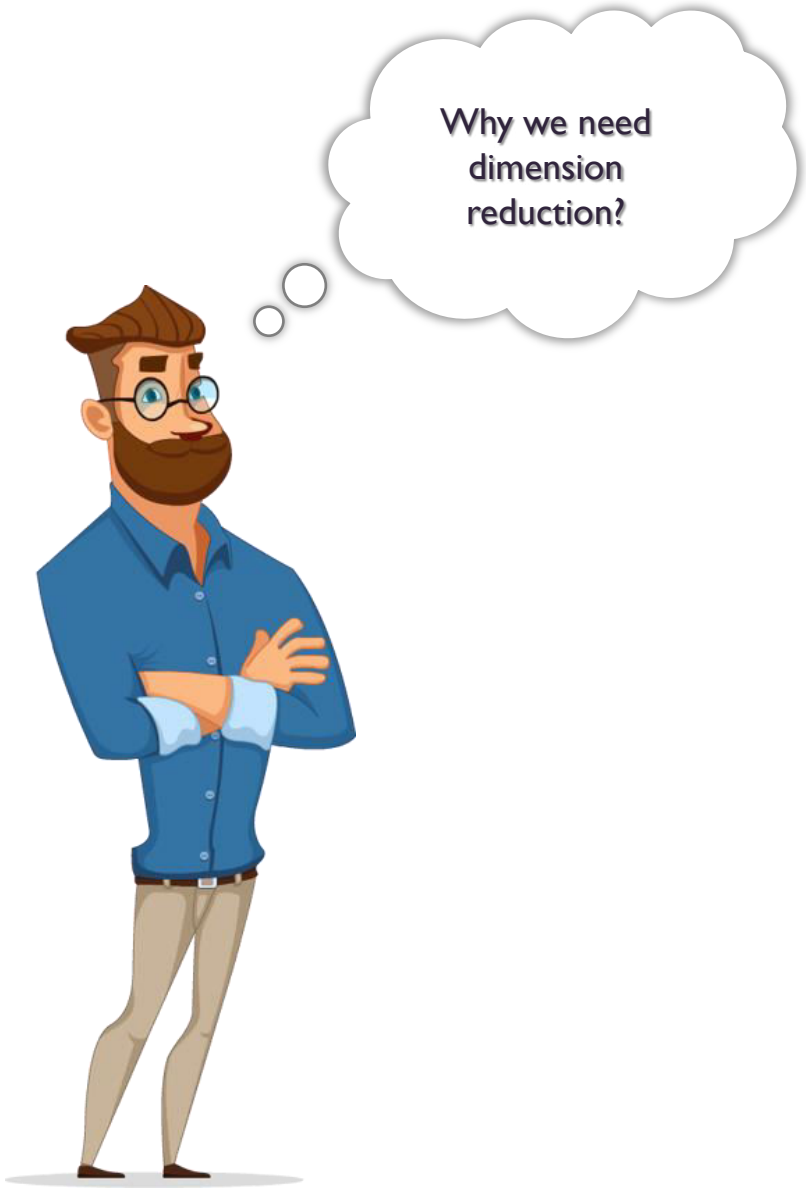


Extracting new variables from old variables

PCA works based on feature extraction

Why dimension Reduction

Dimensionality Reduction



Decrease unwanted dimensions in machine learning

GPS sensors

Gyro meters

Video feeds

Smart devices

Each data will be saved with little incremental information

Data has to be treated to reduce the number of dimensions

Applications of dimensionality Reduction

Dimensionality Reduction

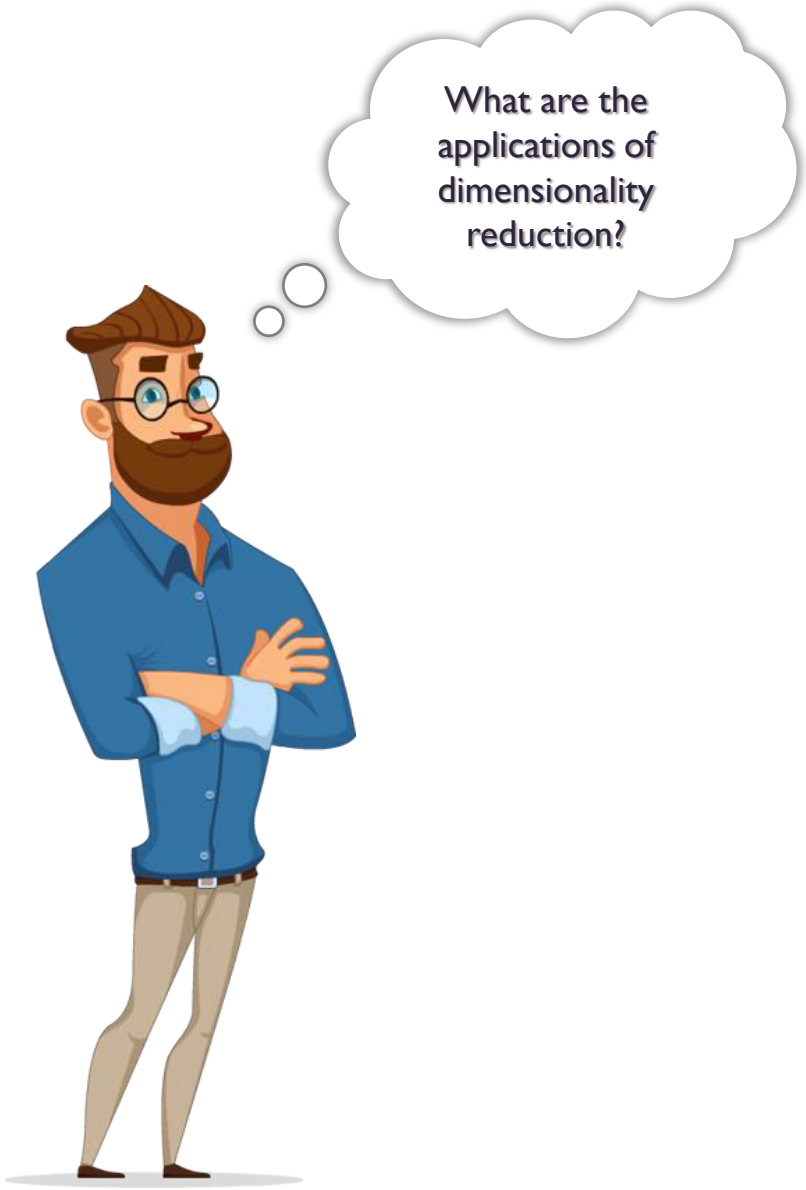


Image Processing

Dimension Reduction

Principal Component Analysis

Principal Component Analysis

Reducing the number of random variables of a given data set

Identify the low-dimension set of axes

Principal Component Analysis

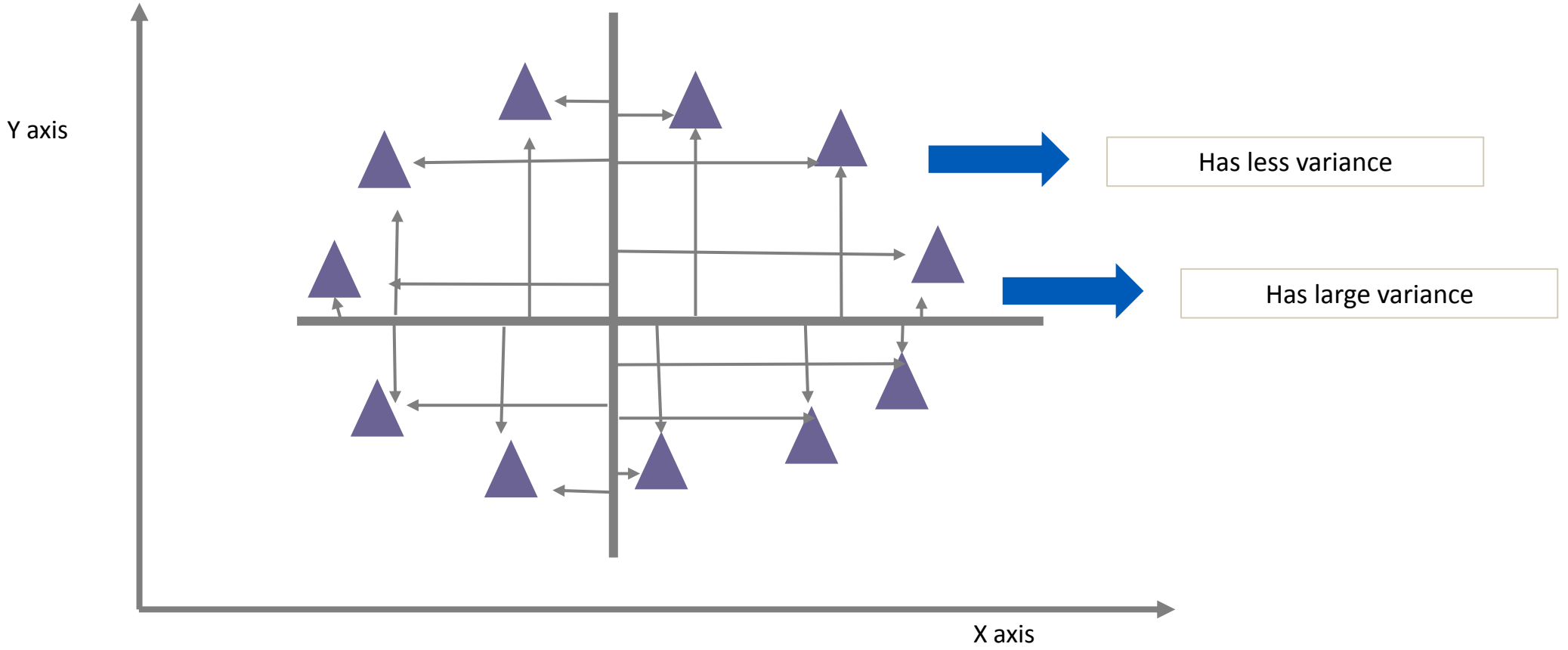
Name	Number of Wheels	Color	Height	Number of seats
Mercedes	4	Red	4 feet	5
BMW	4	Blue	3 feet	5
Marco polo	6	Blue	8 feet	10
Volkswagen	4	White	5 feet	5

Has less variance

Has High variance

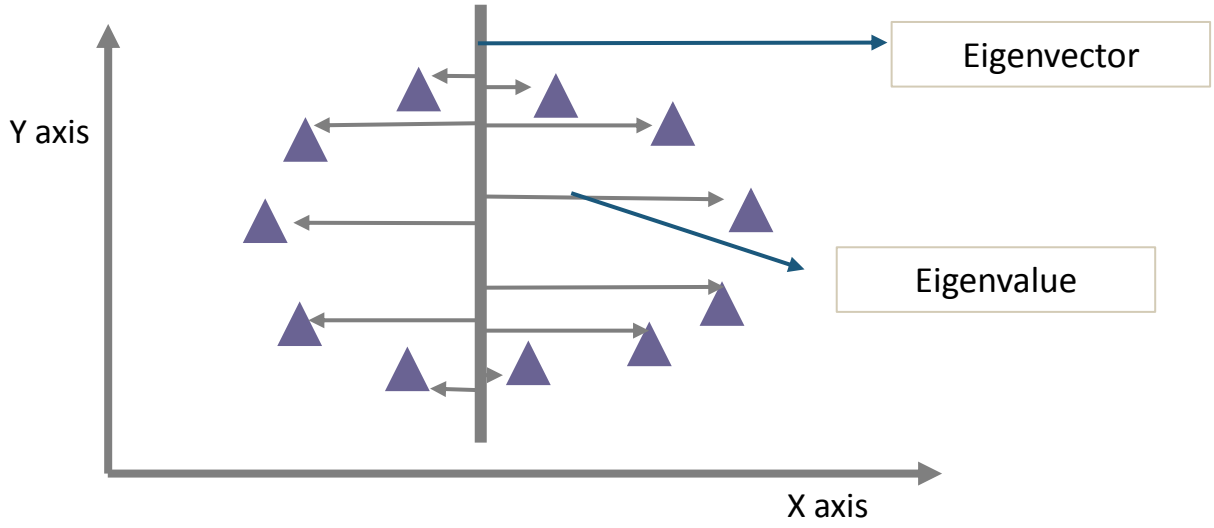
Working mechanism of PCA

Principal Component Analysis

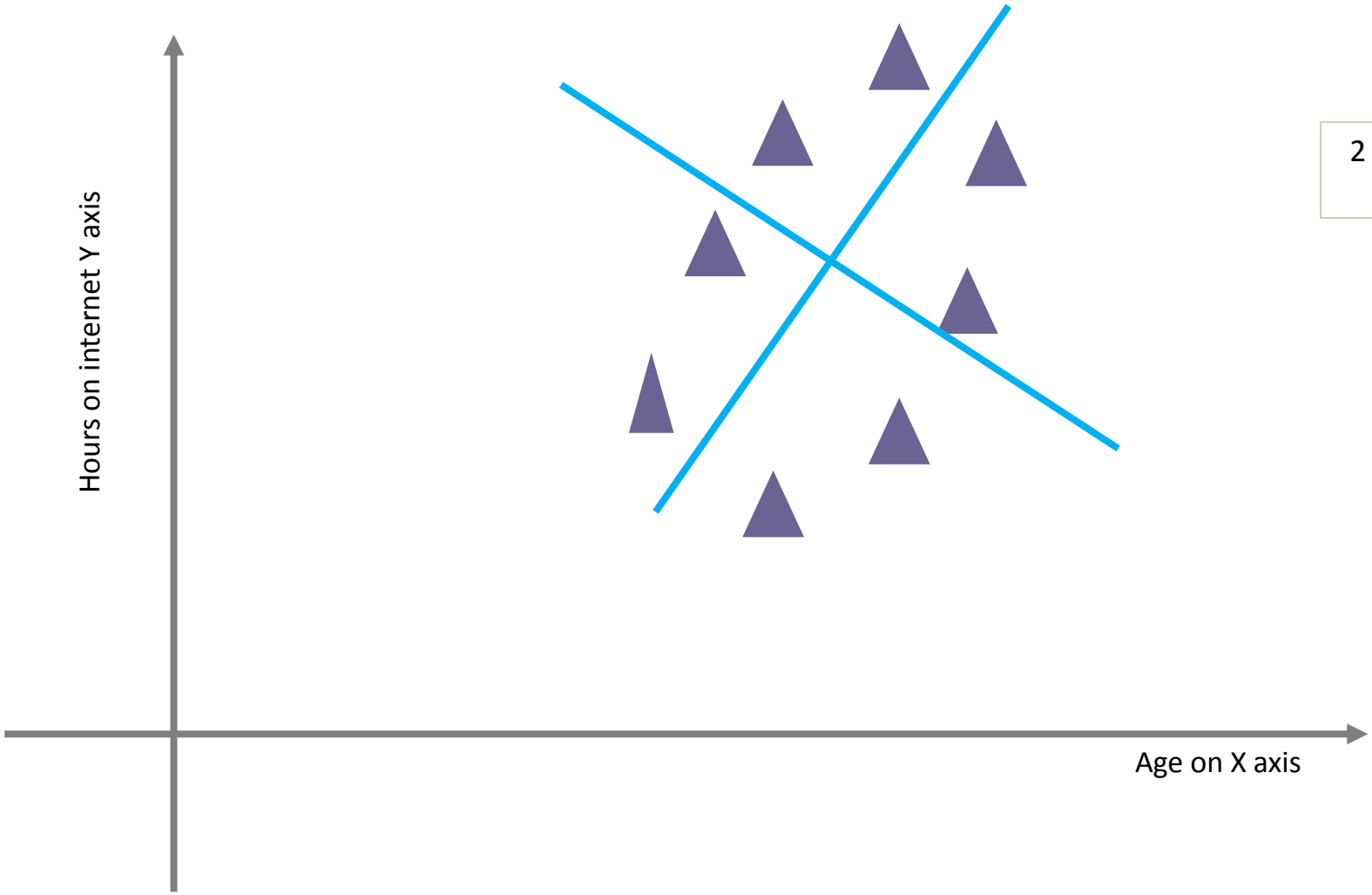


Principal Component Analysis

Eigenvectors and Eigenvalues

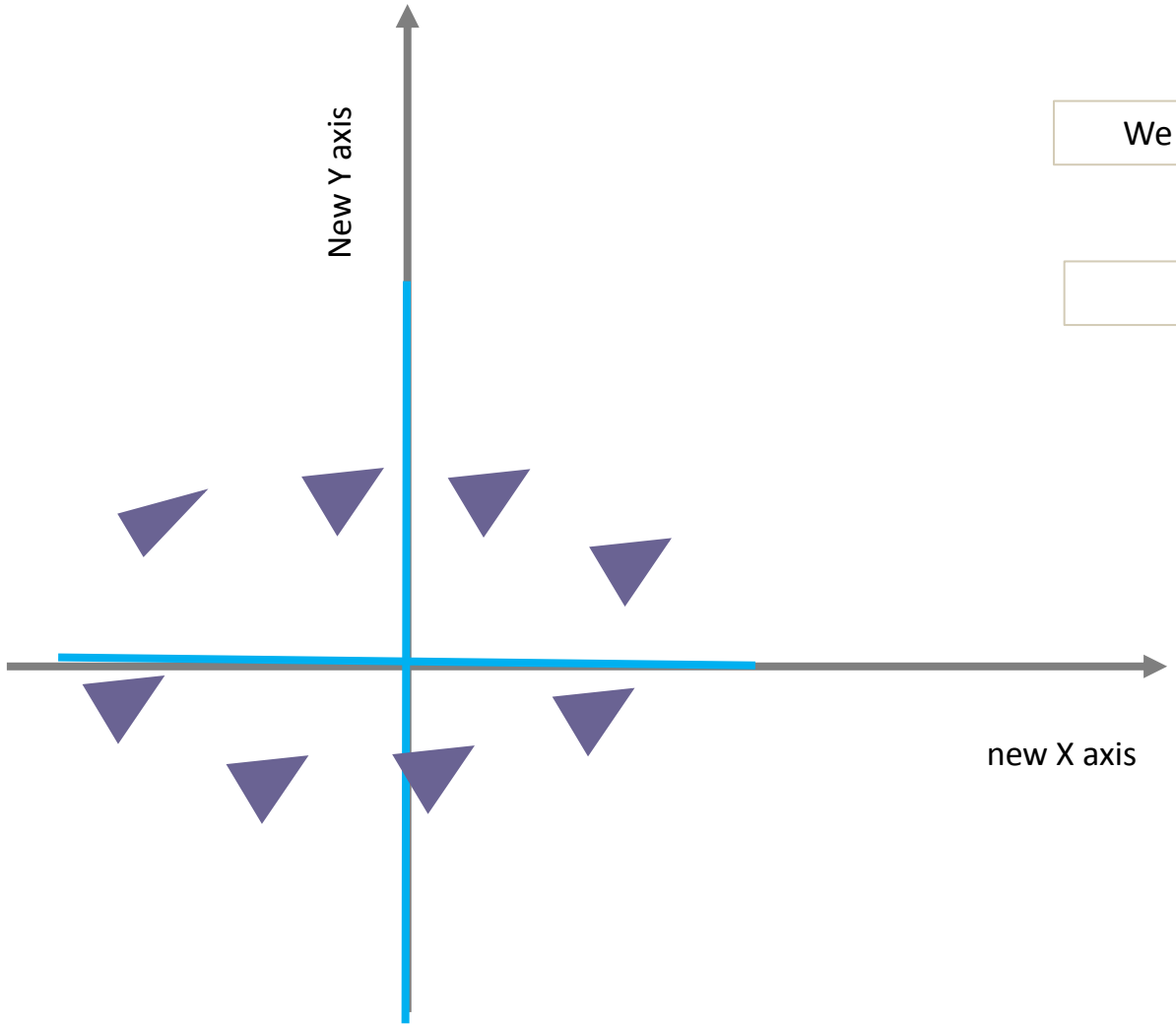


Principal Component Analysis



2 dimension= 2 Eigenvectors and 2 eigenvalues

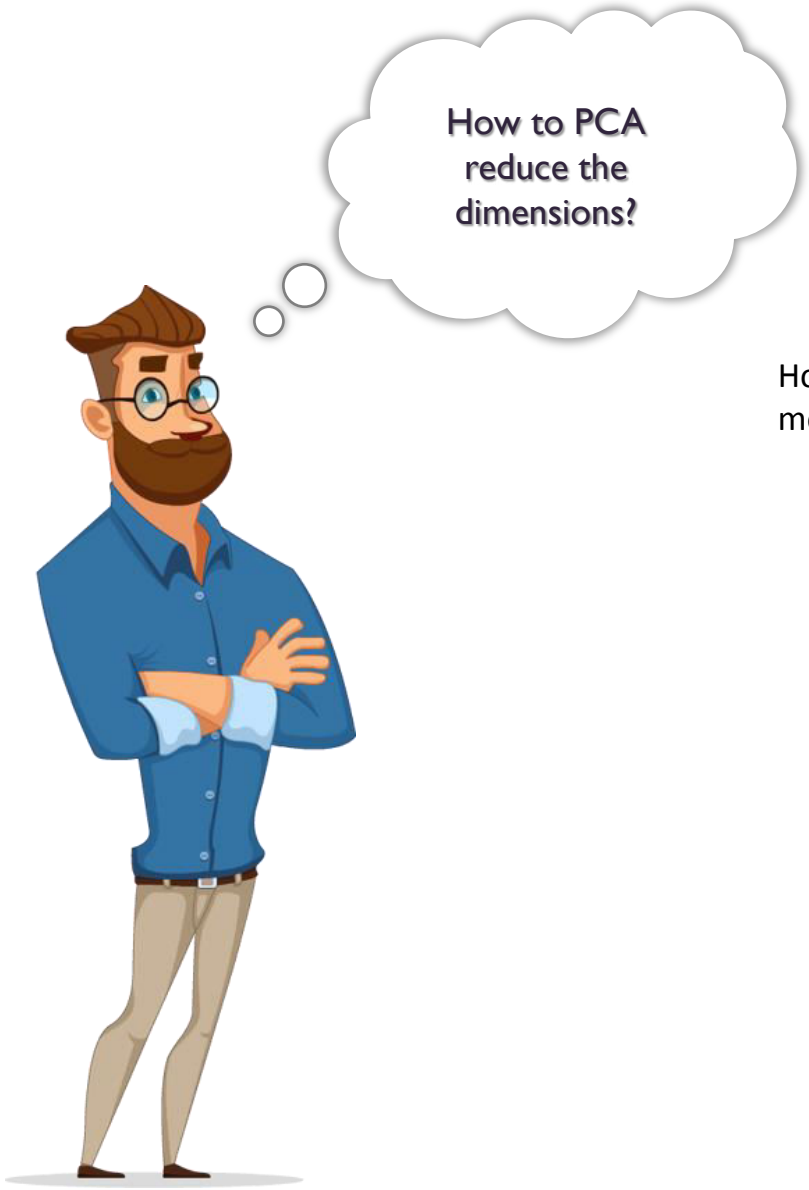
Principal Component Analysis



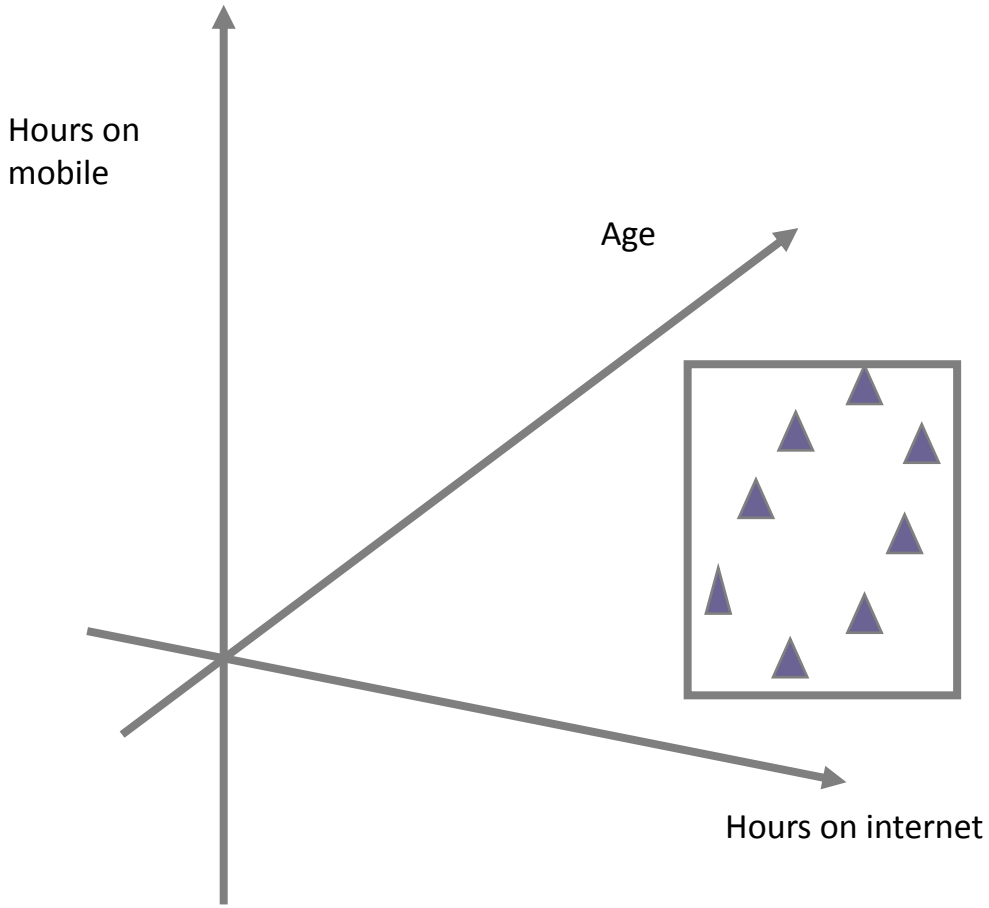
We have just moved the data to origin

More variation

Principal Component Analysis

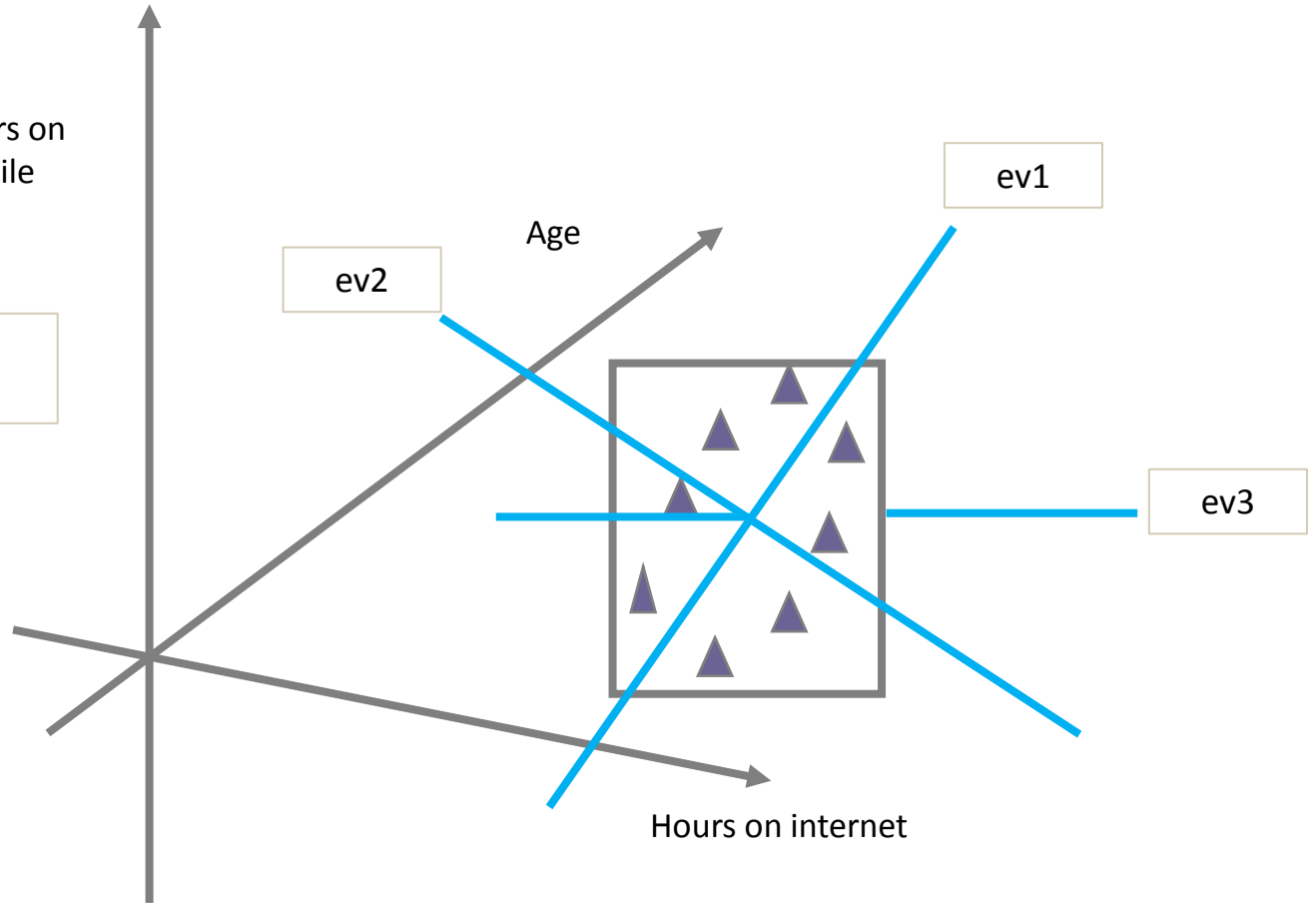


Reduces the data by removing unnecessary dimensions

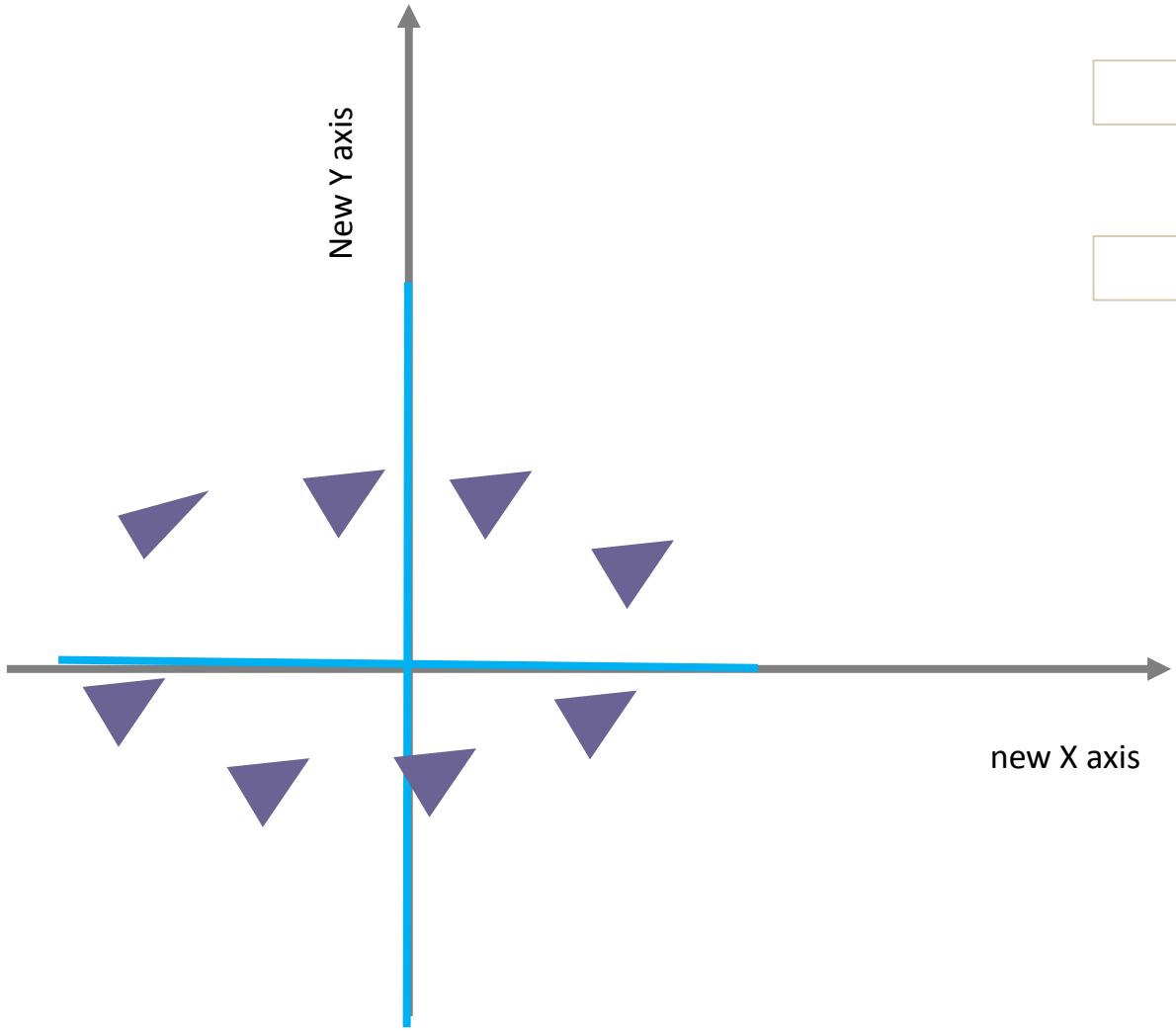


Principal Component Analysis

3 dimensions= 3 eigenvectors and eigenvalues



Principal Component Analysis



3D is reduced to 2D problem

Simplify the data

Thank You