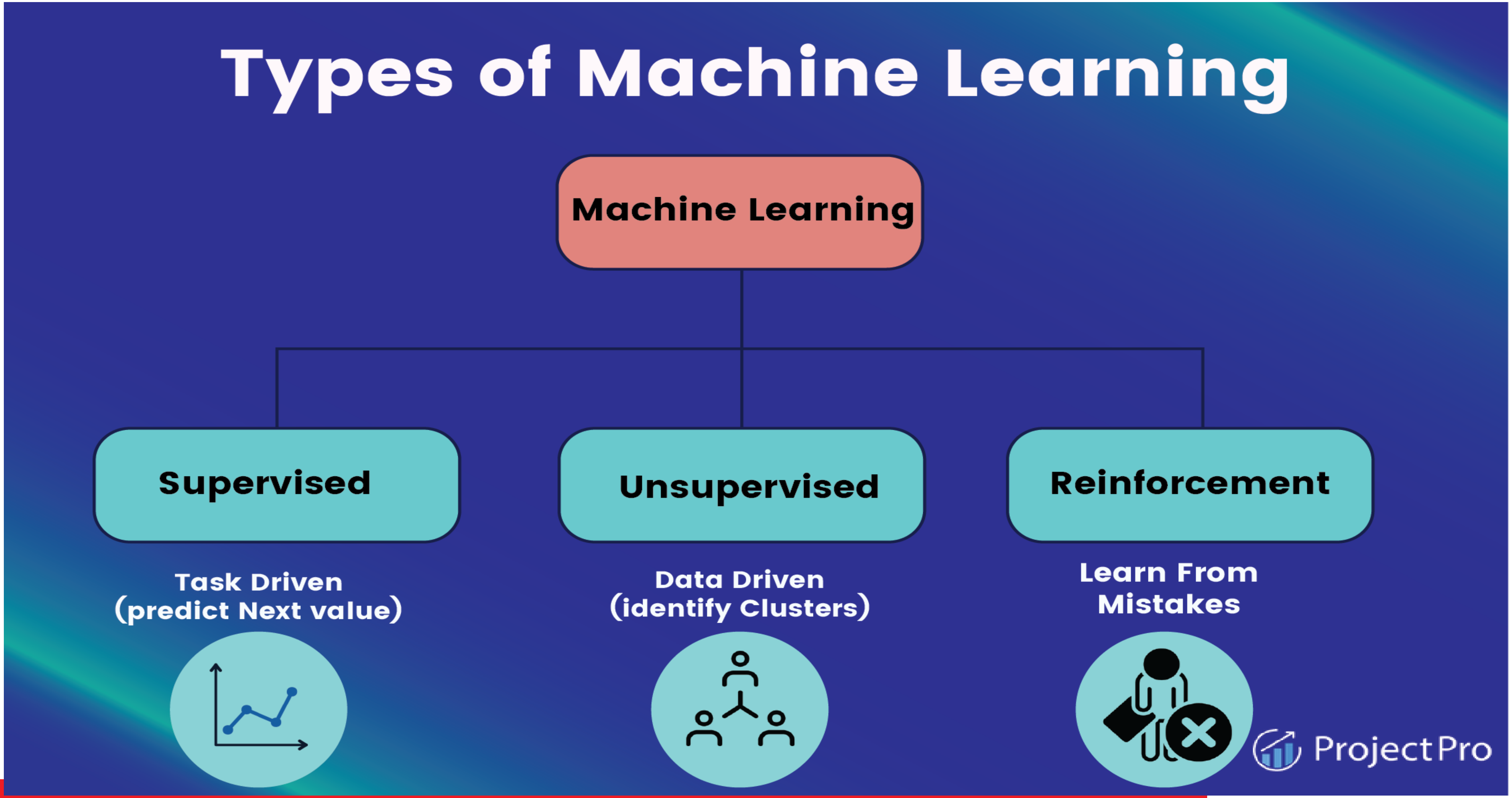
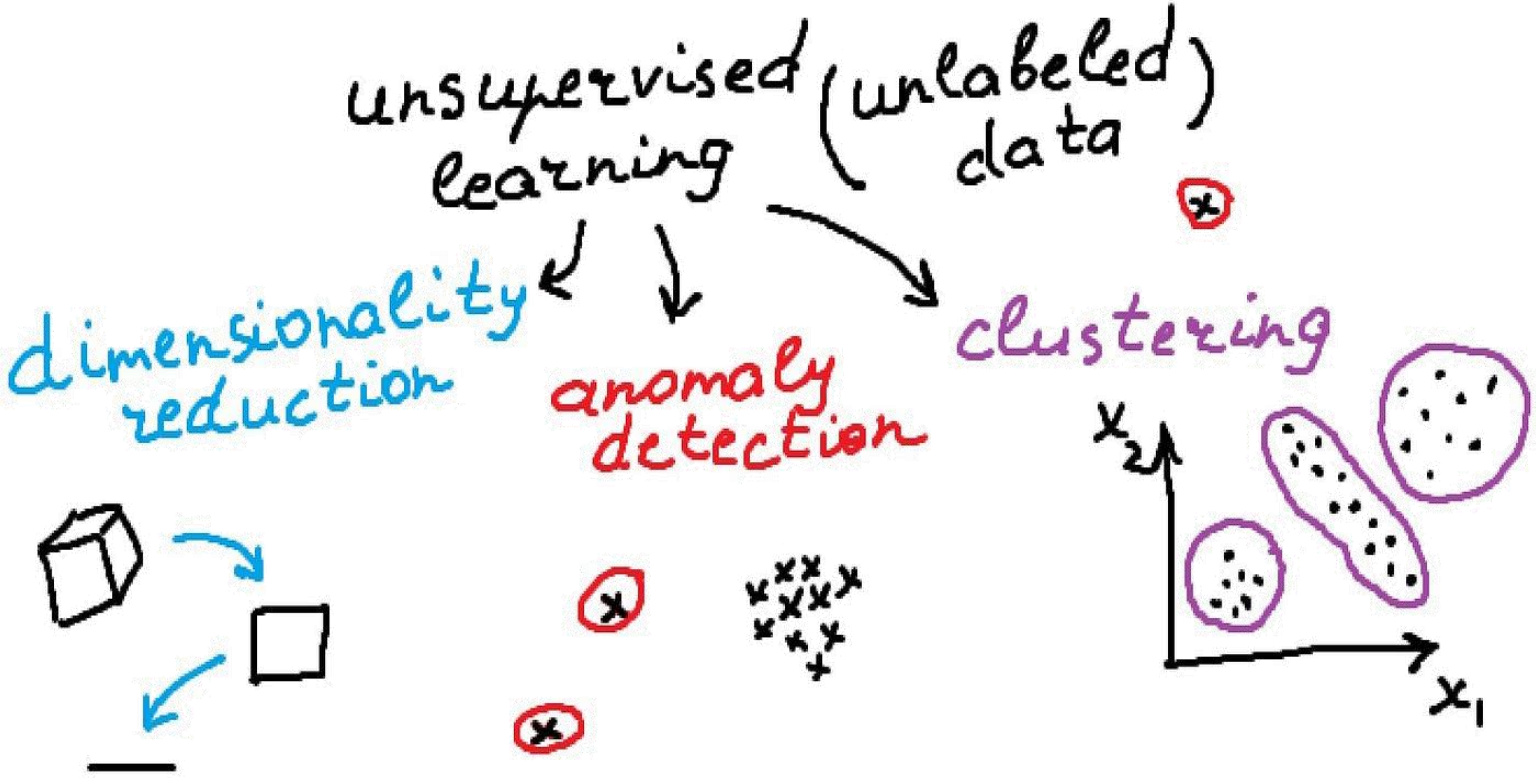
Dimensionality Reduction Techniques

Anushri Tambe

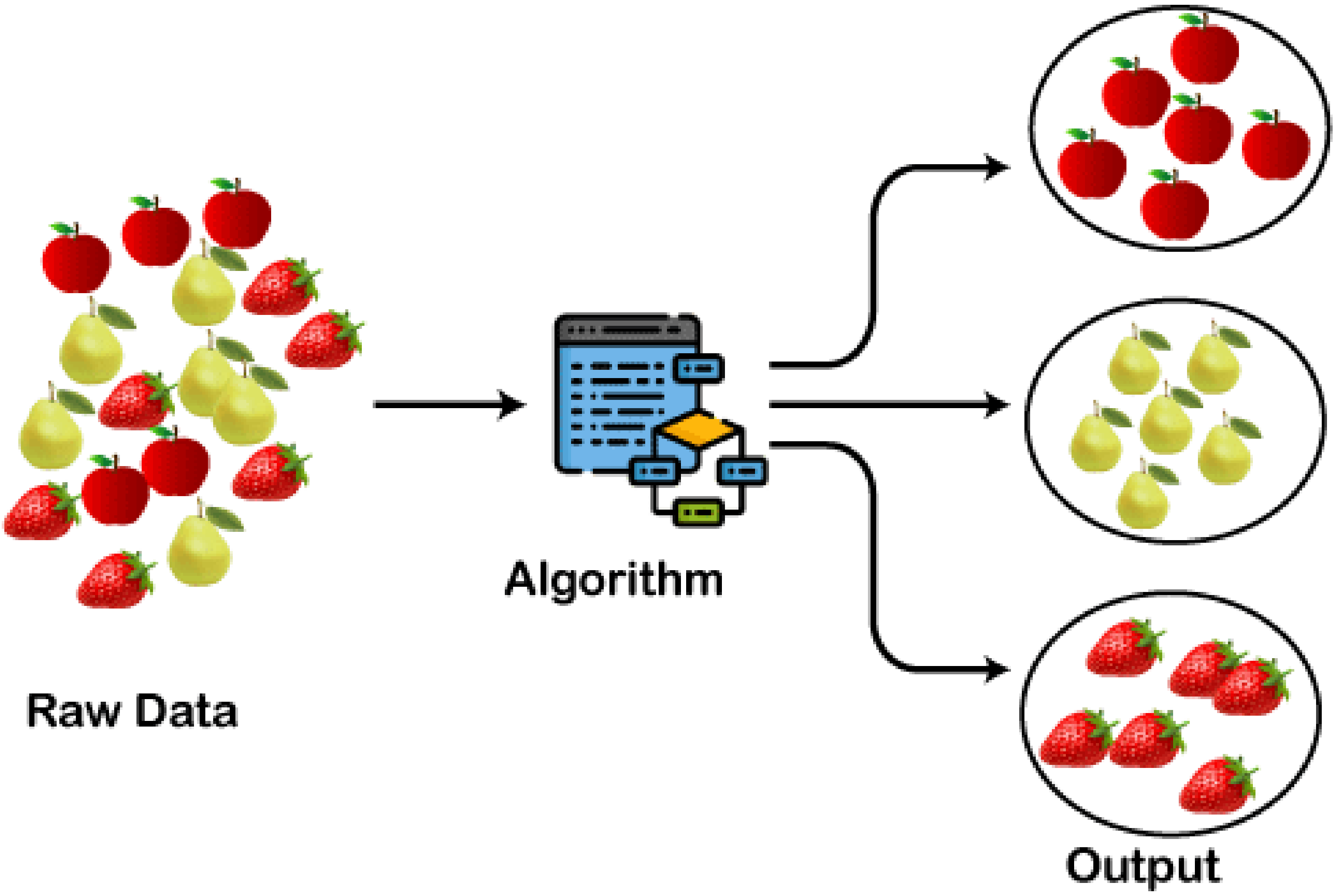

Types of Machine Learning
Machine Learning
Supervised
Unsupervised
Reinforcement
Task Driven (predict Next value)
Data Driven (identify Clusters)
Learn From Mistakes
ProjectPro

Unsupervised learning

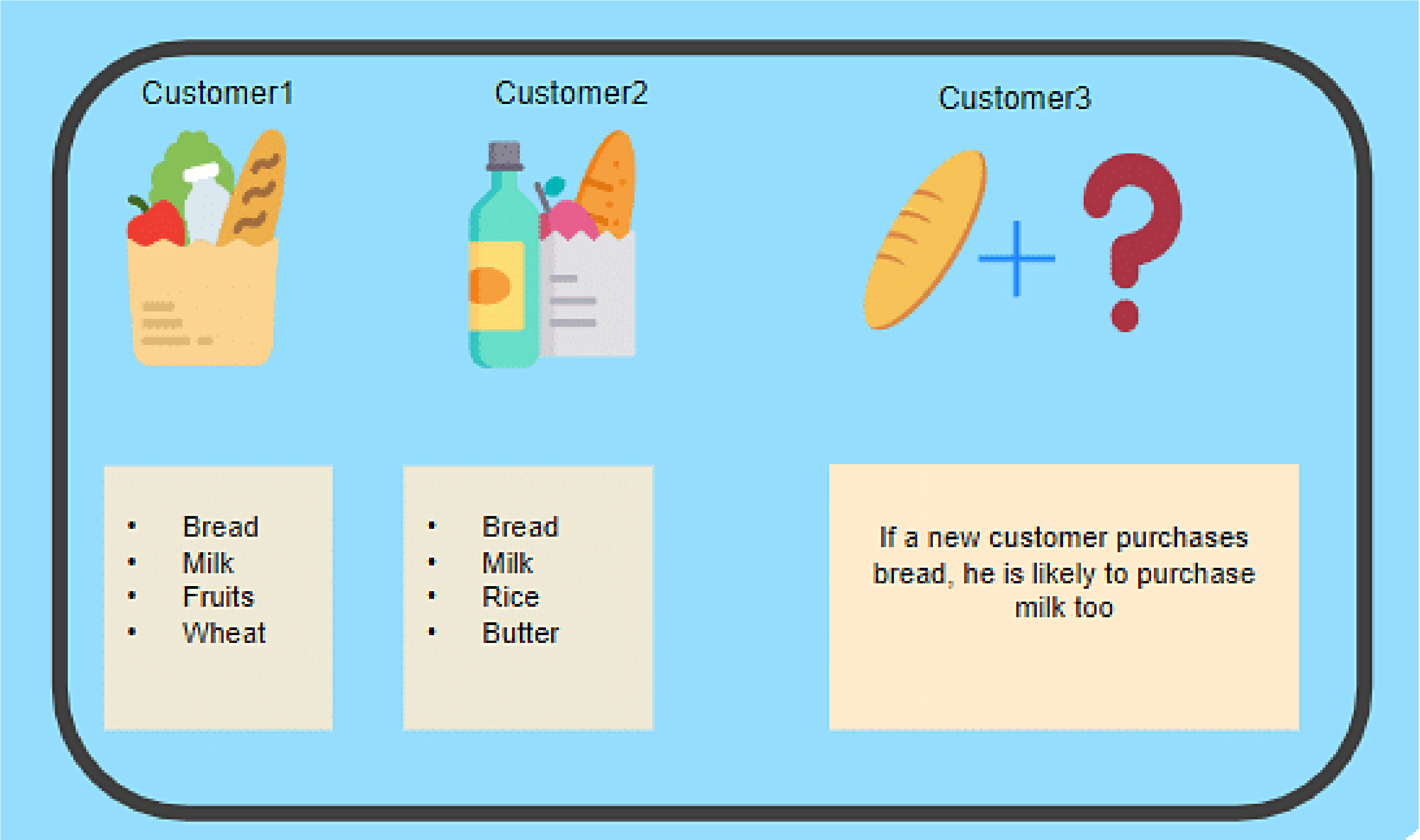
- Analyze and cluster unlabeled datasets
- Discover hidden patterns or data groupings
- Discover similarities and differences in information

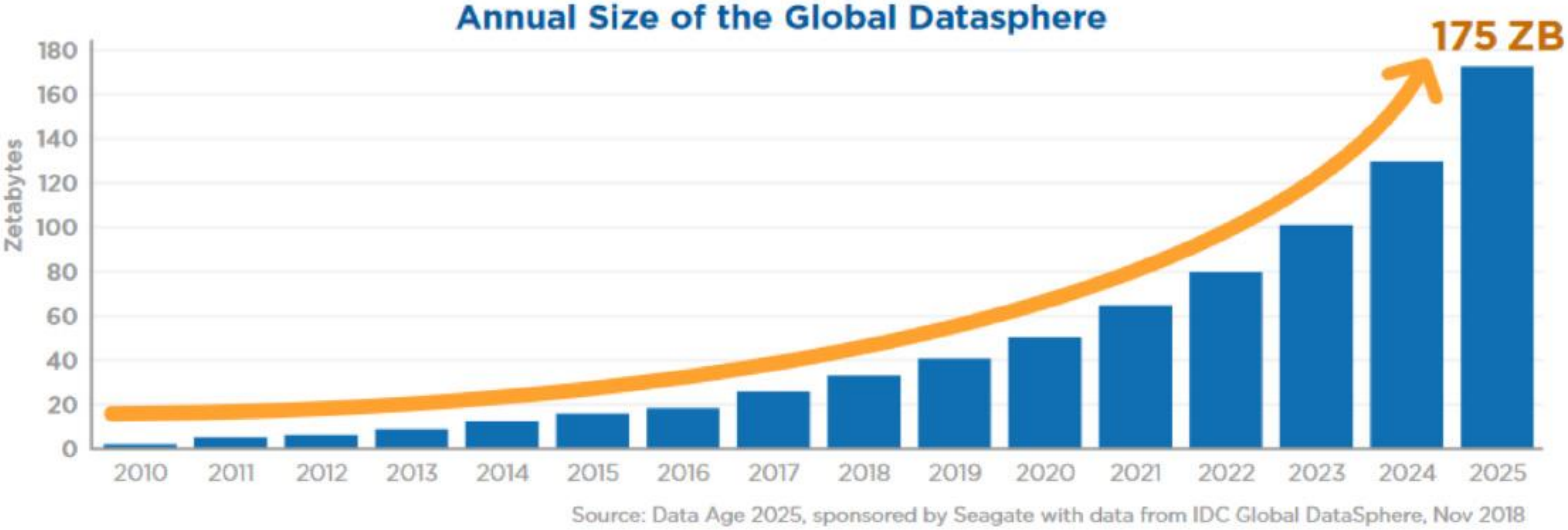


Clustering

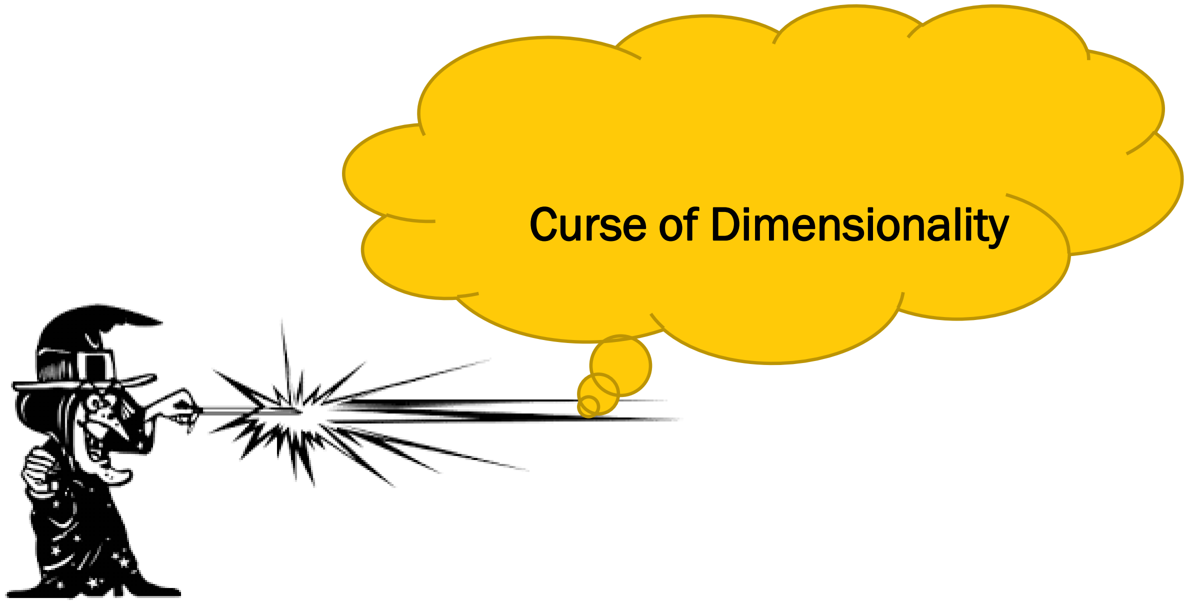


Association



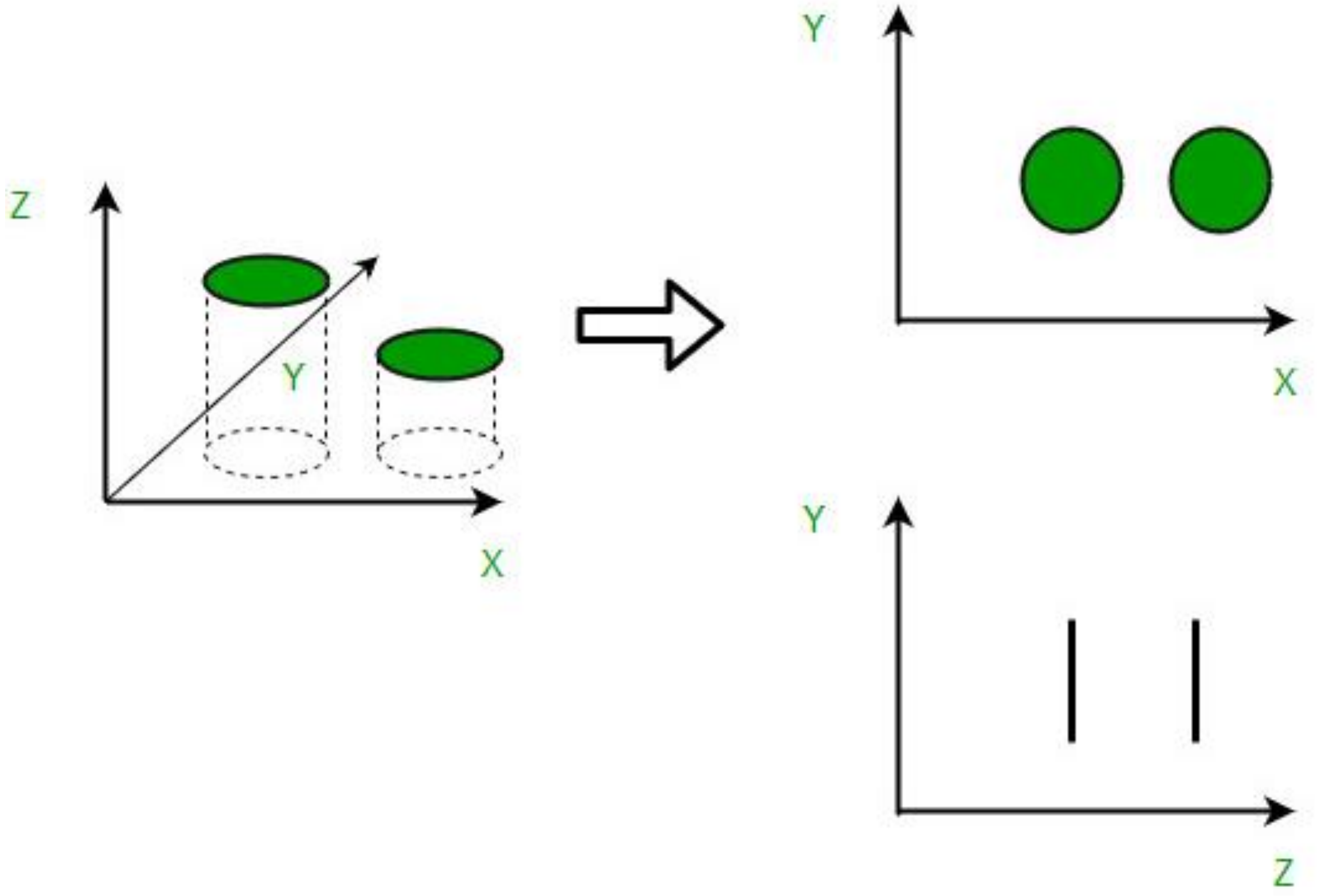

Annual Size of the Global Datasphere
175 ZB
Zetabytes
180
160
140
120
100
80
60
40
20
0
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
Source: Data Age 2025, sponsored by Seagate with data from IDC Global DataSphere, Nov 2018

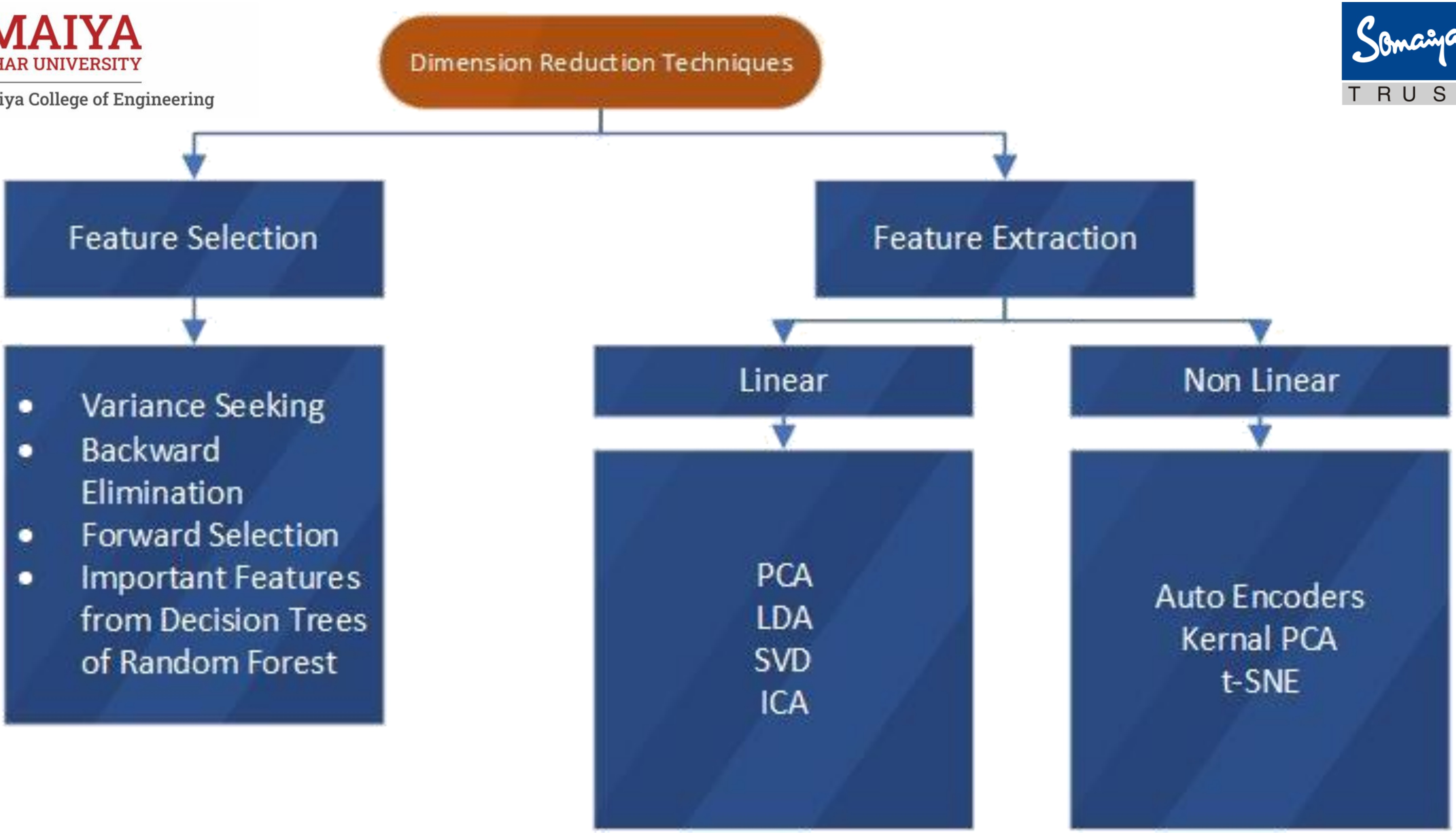
Why we need this?



- Space
- Computation and Training Time
- Multicollinearity
- Visualization
- Noise and Redundancy
- Overfitting

Dimensionality Reduction

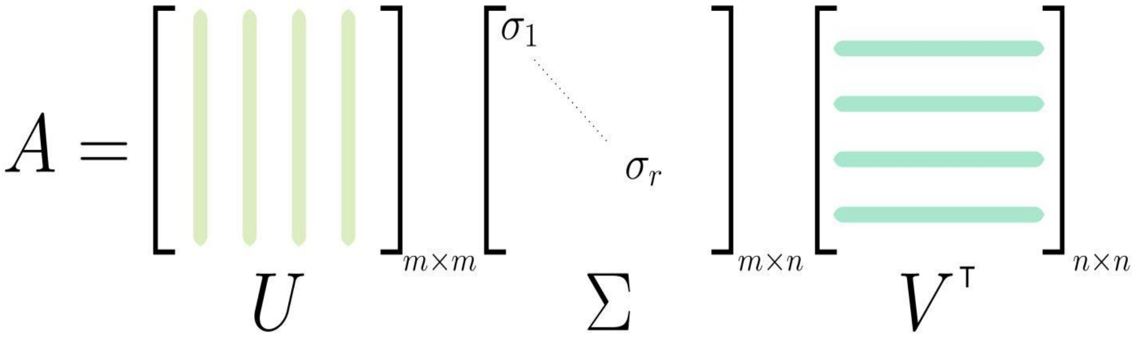



Dimension Reduction Techniques
Feature Selection
Feature Extraction
Variance Seeking
Backward Elimination
Forward Selection
Important Features from Decision Trees of Random Forest
Linear
Non Linear
PCA
LDA
SVD
ICA
Auto Encoders
Kernal PCA
t-SNE

What is SVD?

Any $m \times n$ matrix A can be factored into the product of three matrices as follows

$$A = U\,\Sigma\,V^T$$



Here,

U is $m \times m$ orthogonal matrix

Σ is $m \times n$ diagonal matrix

V is $n \times n$ orthogonal matrix

$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T$$

Steps to find SVD $A = U \Sigma V^T$

Columns of U are orthonormal Eigen vectors of AA^T and Columns of V are orthonormal Eigen vectors of A^TA. Diagonal entries of Σ are square root of Eigen values of AA^T and A^TA (as they are same) arranged in descending order.

We can use following steps as well.

1. Find A^TA.
2. Find Eigen values and Eigen vectors of A^TA.
3. Arrange all eigen values of A^TA in descending order say $\lambda_1 > \lambda_2 > \lambda_3 \ldots \ldots > \lambda_r$
   $\sigma_i = \sqrt{\lambda_i}$. Construct matrix Σ by writing σ_i in the diagonals.
4. Make sure that Eigen vectors of A^TA are orthonormal vectors.
   These orthonormal Eigen vectors of A^TA are columns of V .
5. $u_i = \frac{1}{\sigma_i} Av_i$. Make sure that these u_i's are also orthonormal vectors.
   These orthonormal vectors u_i are columns of U.
6. Write $A = U \Sigma V^T$

Problem: Construct SVD of $\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & -1 \end{bmatrix}$

Using SVD we can decompose A as $A_{3\times 2} = U_{3\times 3}\,\Sigma_{3\times 2}\,V^T_{2\times 2}$

1. $A^TA = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$
2. Characteristic equation of A^TA is $\lambda^2 - 6\lambda + 8 = 0$

   $\therefore$ Eigen values of A^TA are $\lambda = 4, 2$

   To find eigen vector $v \neq 0$ we need to solve $(A - \lambda I)v = 0$

   For $\lambda = 4$ we get $(A - 4I)v_1 = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = 0$

   $\Rightarrow x - y = 0 \Rightarrow x = y$ Hence, for $x = y = 1$, We get $v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

   For $\lambda = 2$ we get $(A - 2I)v_1 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = 0$

   $\Rightarrow x + y = 0 \Rightarrow y = -x$ Hence, for $x = 1, y = -1$

   We get $v_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

   Corresponding Eigen Vectors are $v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, v_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$
3. Clearly $\lambda_1 = 4 > \lambda_2 = 2 \therefore \sigma_1 = \sqrt{\lambda_1} = 2$ and $\sigma_2 = \sqrt{\lambda_2} = \sqrt{2}$

   As Σ is diagonal matrix of order of 3×2

   $$\Sigma = \begin{bmatrix} 2 & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \end{bmatrix}$$
4. Clearly, v_1 and v_2 are orthogonal but they are not unit vectors.

   $$\frac{v_1}{\|v_1\|} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}, \frac{v_2}{\|v_2\|} = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$$

   $$V = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

5. As $u_i = \frac{1}{\sigma_i} A v_i$

$$u_1 = \frac{1}{\sigma_1} A v_1 = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 2/\sqrt{2} \\ 2/\sqrt{2} \\ 0 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}$$

$$u_2 = \frac{1}{\sigma_2} A v_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 0 \\ 0 \\ 2/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Clearly, u_1 and u_2 are orthogonal vectors.

Also, $\|u_1\| = 1, \|u_2\| = 1$

As U is 3×3 matrix let us find u_3.

Also, we want $u_3 = (a, b, c)$ as orthogonal to u_1 and u_2

$\therefore u_3.u_1 = 0$ and $u_3.u_2 = 0$

Hence, $(a, b, c) \cdot \left(1/\sqrt{2}, 1/\sqrt{2}, 0\right) = a/\sqrt{2} + b/\sqrt{2} + 0 = 0$

$\therefore a = -b$

Also, $(a, b, c) \cdot (0,0,1) = 0 + 0 + c = 0$

$\therefore c = 0$

Putting $b = 1$, we get $a = -1$. Hence, $u_3 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$

But, $\|u_3\| = \sqrt{2}$

$$\therefore \frac{u_3}{\|u_3\|} = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix} \quad \text{Hence, } U = \begin{bmatrix} 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1 & 0 \end{bmatrix}$$

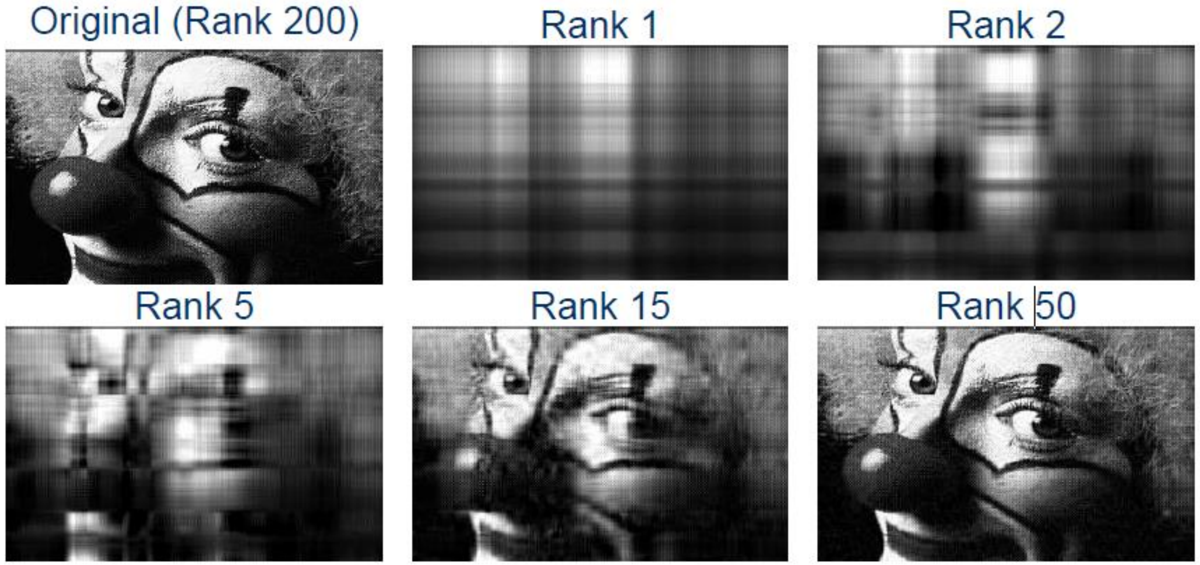
Result:

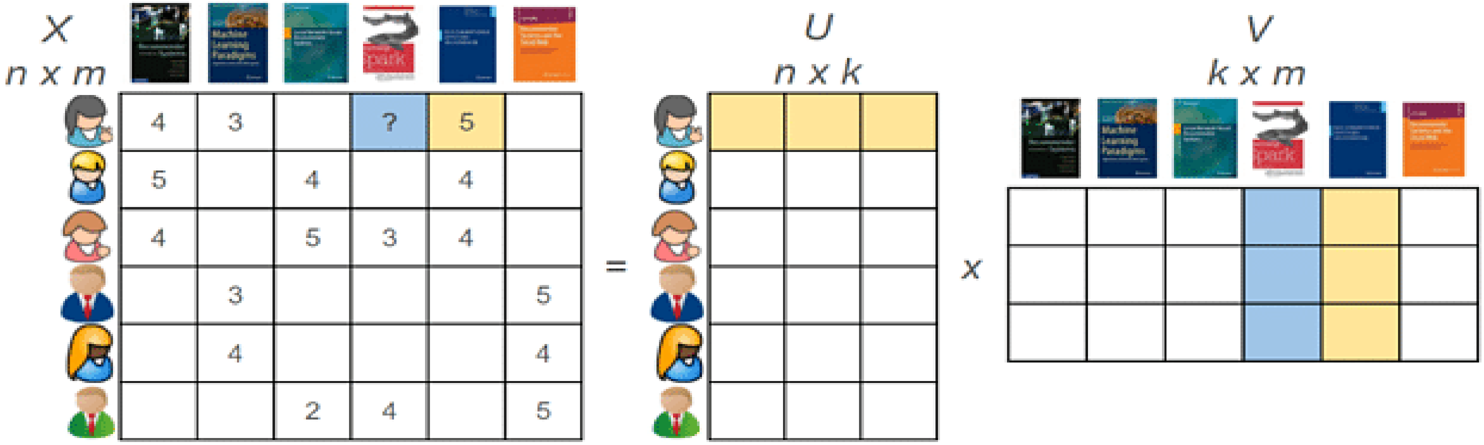
Singular Value Decomposition of A is

$$A = U\,\Sigma\,V^T = \begin{bmatrix} 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

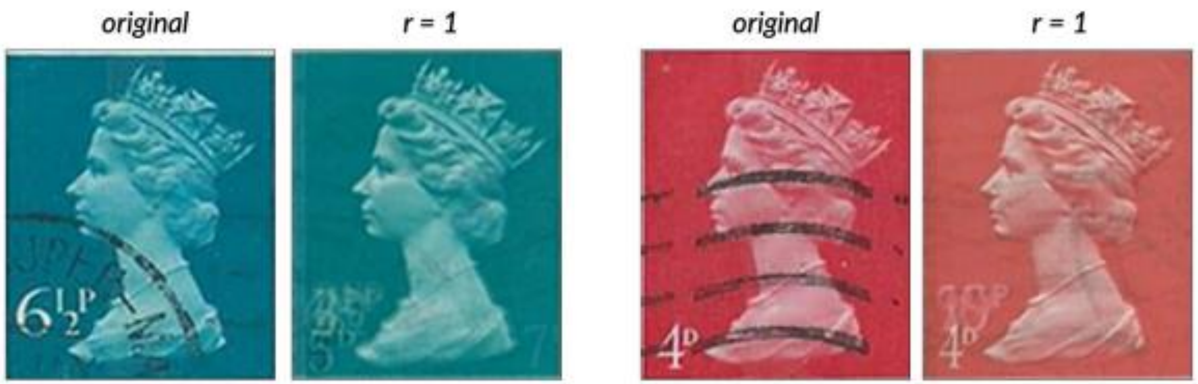
Applications

- Image Compression
- Recommender System
- Image Denoising



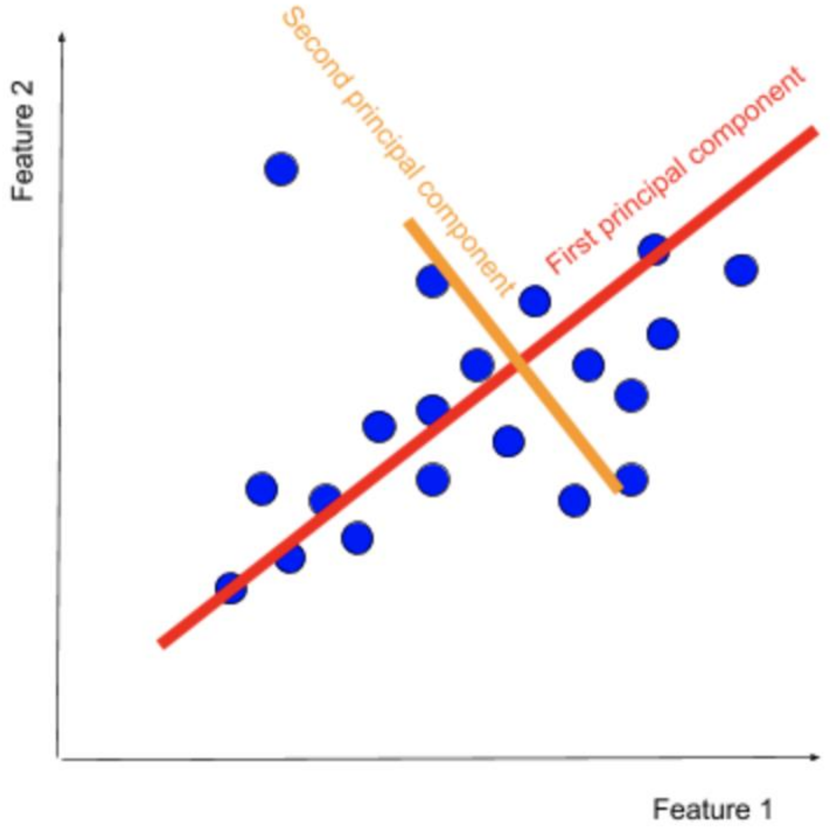


$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T$$



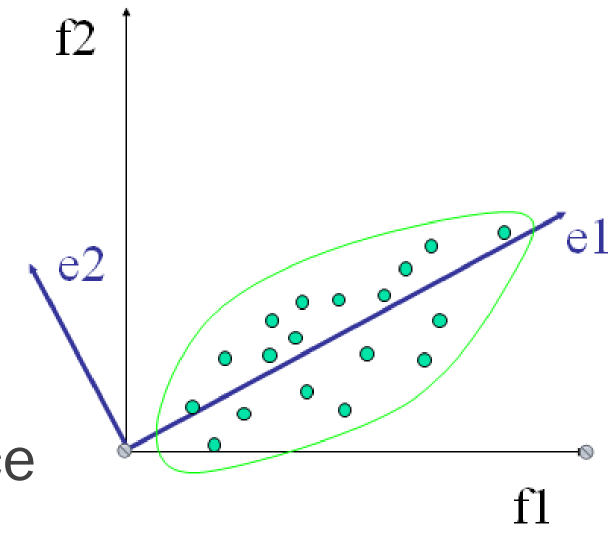
Principal Component Analysis

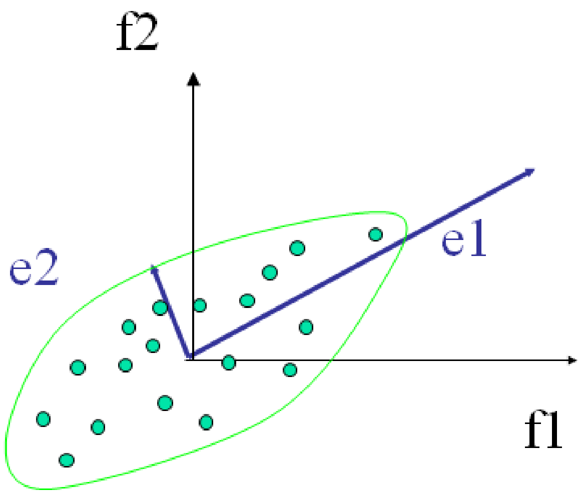
Variables are transformed into a new set of variables which are linear combination of original variables.



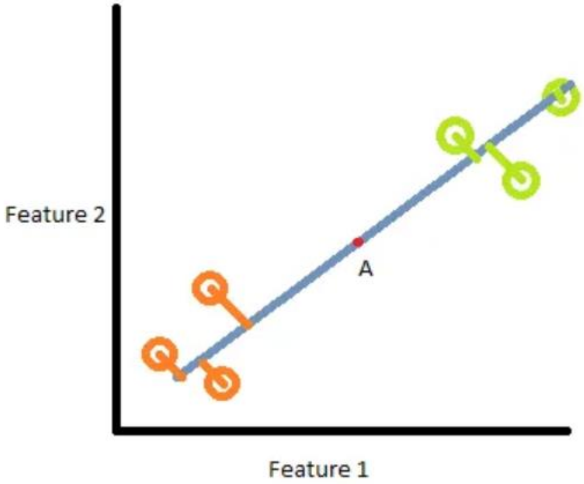
How it works?

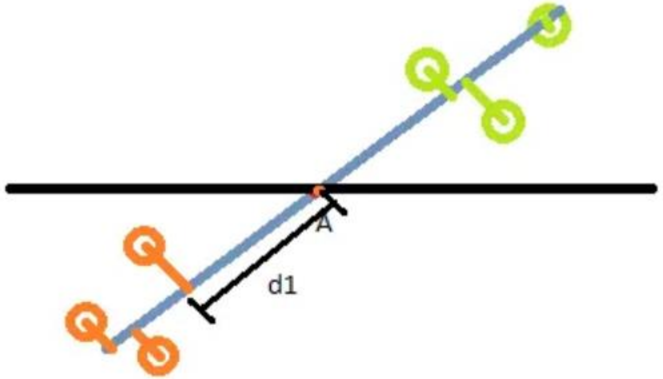
- Variance
- linear combinations of the features
- Principal Components
- Select Components with high variance

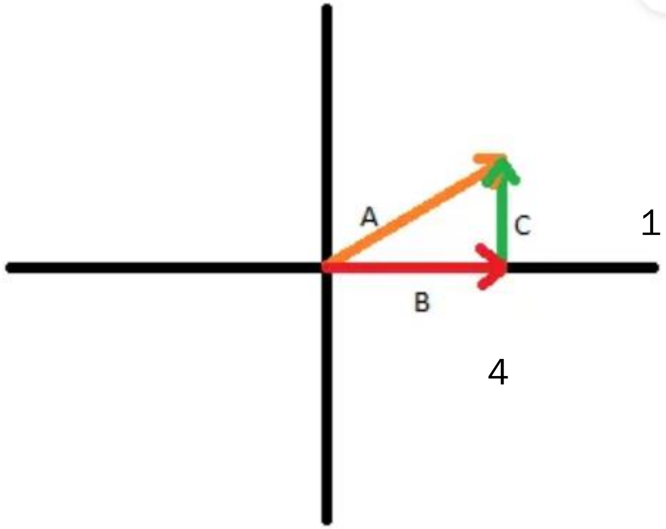


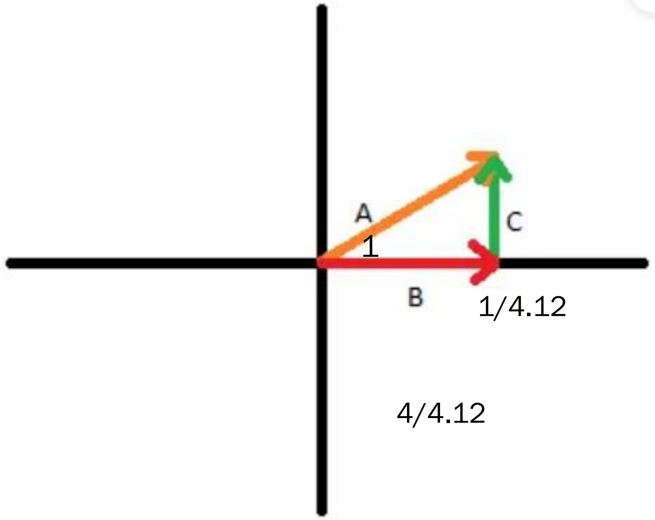


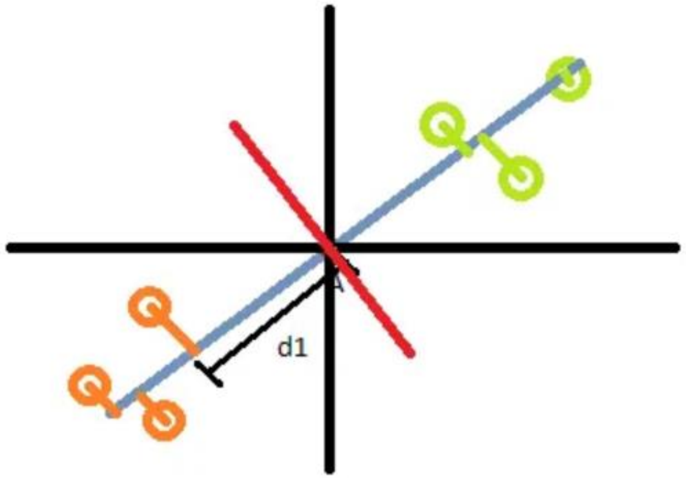
PCA is orthogonal projection or transformation of the data into a subspace such that we get max variance











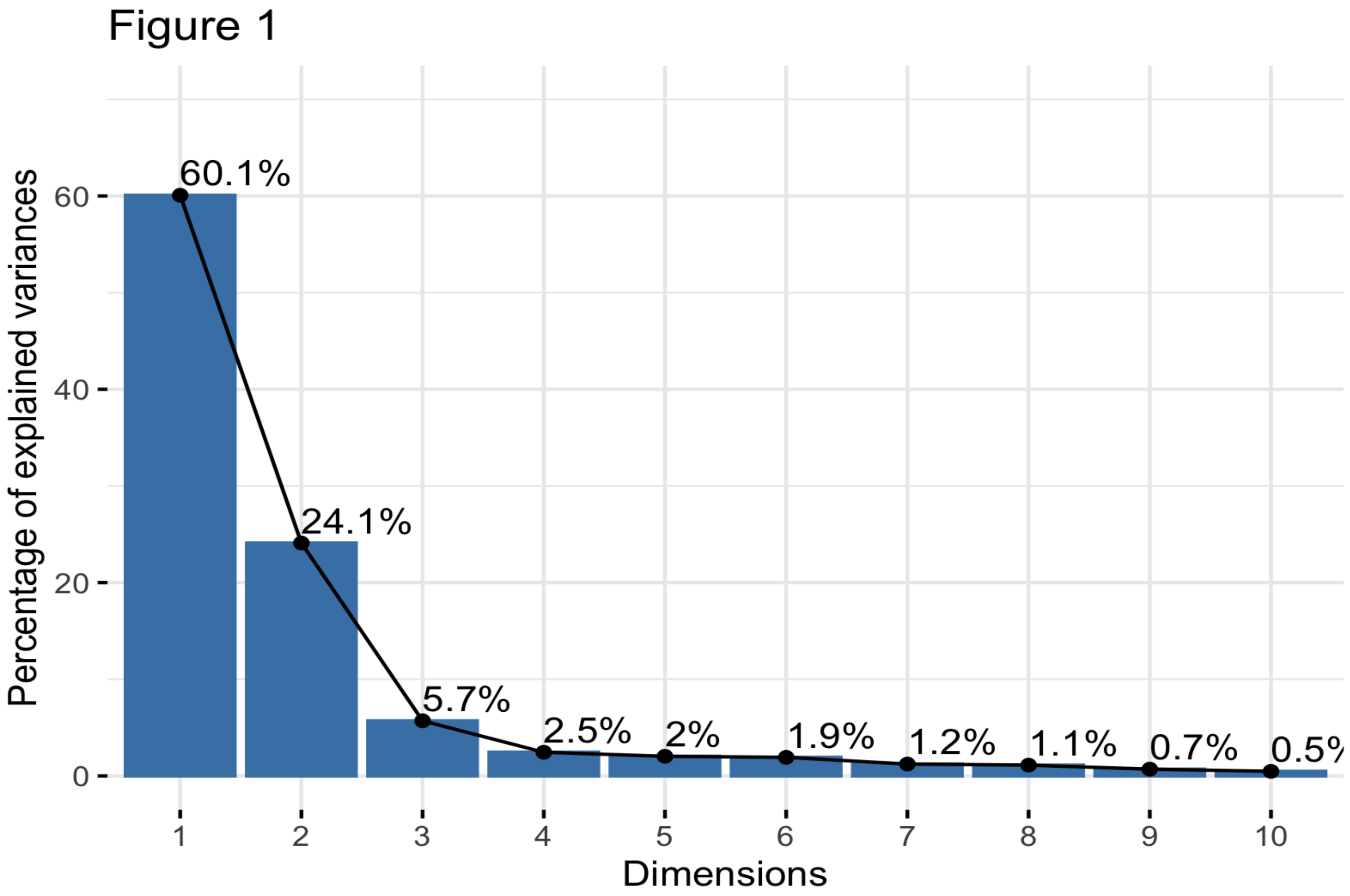
The Sum of Squared Distances d1,d2,d3,d4,d5,d6 is the eigenvalue and A is eigen vector.

Steps to perform PCA

- Find Covariance Matrix
- Find its Eigen Values and Eigen Vectors
- Find the eigenvectors with the largest eigenvalues correspond to the dimensions that have the strongest correlation in the dataset

VIDYAVIHAR UNIVERSITY
K J Somaiya College of Engineering

Scree Plot




рахмат
danke
謝謝
ngiyabonga
شكراً جزيلاً
spas
Баярлалаа
mersi
kia ora
barka
welalin
tack
teşekkür ederim
спасибо
faafetai lava
vinaka
спасибі
blagodaram
dank je
misaotra
matondo
paldies
grazzi
mahalo
tapadh leat
thank you
gracias
хвала
asante
manana
obrigada
tenki
kiitos
dankie
nanni
nandri
dhanyavad
hvala
maururu
köszönöm
aciū
dankon
akun
bedankt
bayarlalaa
gràcie
enkosi
dziękuję
djiere dieuf
tau
mochchakkeram
murakoze
дякую
mamnun
chnorakaloutioun
gratias ago
gràcies
sulpáy
go raibh maith agat
chokrane
sobodi
děkuji
taiku
sukriya
kop khun krap
sagolun
obrigado
arigatô
takk
dakujem
trugarez
mèsi
ありがとう
tanemirt
grazie
didi madloba
najis tuke
terima kasih
rahmet
diolch
dhanyavadagalu
shukriya
merce
мерси
kam sah hamnida
rahmat
ευχαριστώ
তোমাকে ধন্যবাদ
감사합니다
xièxie
merci