Principal Component Analysis (PCA) [Dimensionality Reduction]

⇒ Price of the House ⇐

1. House size
2. No.of bedrooms
3. No.of bathrooms

① Curse of Dimensionality.

Dataset = 500 features ⇒

3 features → M1 → Acc 1

6 features → M2 → Acc2 ↑↑

15 features → M3 → Acc 3 ↑↑

50 features → M4 → Model is overfeeded } → Acc ↓↓

Acc 1 < Acc2 ↑↑ < Acc 3 ↑↑

100 features → M5 → Acc5 ↓↓

M6.

House Price

Person ↑ Confused

↓↓ Price Rate

Acc ↓↓.

Mumbai

1bhk

→ Loc A ← 25-35 lakhs

3bhk ← 80-1cr

beach ← 2-5cr

Sharukh khan ← 5-6 cr

Grocery shop

Schools

- .

f1 f2 . . . f100

↓↓

PCA

f1 f2 - - - f20

Two different ways to remove Curse of Dimensionality

500 D
⇓
20 D ⇒ Problem

① Feature Selection
⇓
Imp features

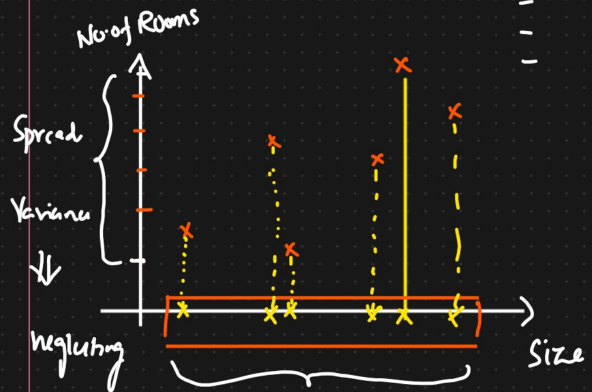
② Dimensionality Reduction (PCA)
⇓
Feature Extraction
⇓
Principal Component Analysis
⇓
Eigen Value & Vectors

[500 features]
⇓
[20 features]

PCA Geometric Intuition

[Dimensionality Reduction]

x_1 Size of Rooms	x_2 No. of Rooms	Price ↗ O/P
-	-	-



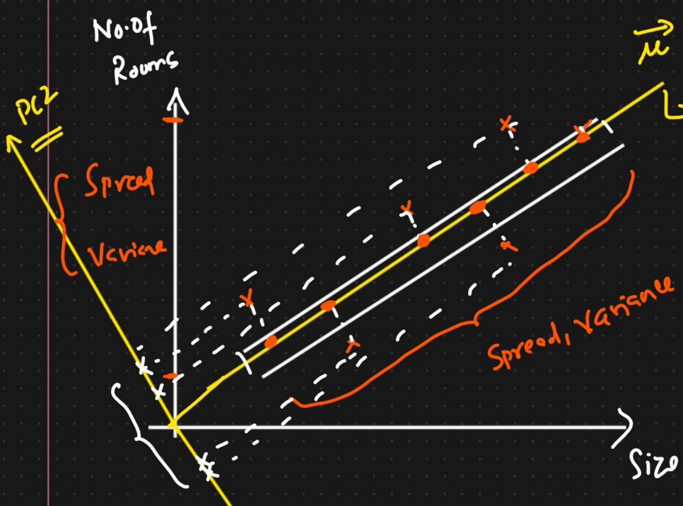
Aim → 2d → 1 dimension

① Feature Selection
⇓
2 D → 1 D

3D-2D
PC1, PC2 | PC3

Feature Extraction
⇑
② PCA → 2D → 1D

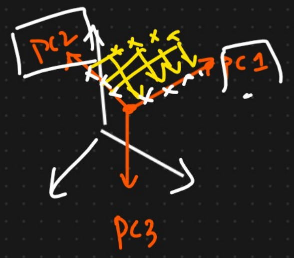
Variance, Spread
⇓
Information of the DATA



↳ PC1 {Principal Component 1}

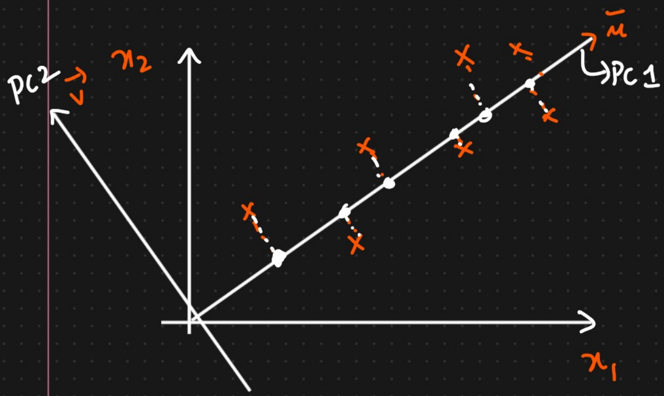
2 features → PC1, PC2

3 feature → PC1, PC2, PC3



3D - 2D

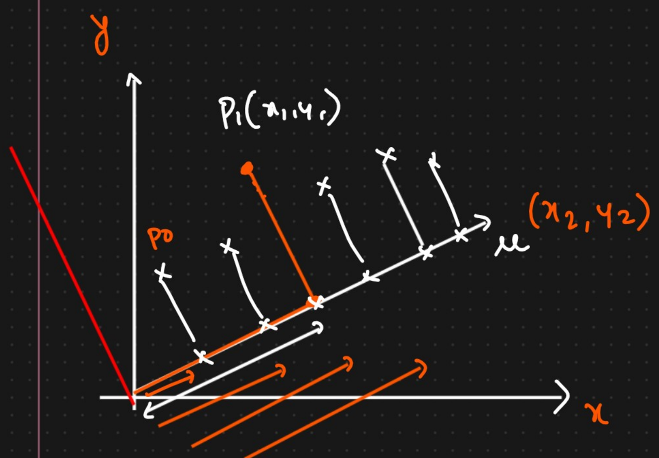
Maths Intuition behind PCA Algorithms



2D → 1D

PC1 > PC2

1. Projections
2. Optimization → Principal Component Max variance



$$\text{Proj}_{P_1} u = \frac{P_1 \cdot u}{\|u\|} \qquad \|u\| = 1$$

$$\boxed{\text{Proj}_{P_1} u = P_1 \cdot u} \qquad \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} \cdot \begin{bmatrix} x_2 & y_2 \end{bmatrix}$$

$$= \boxed{x_1 x_2 + y_1 y_2} \Rightarrow \text{Scaler value}$$

$$\boxed{P_0', P_1', P_2', P_3' \ldots P_n'}$$

⇓

Scaler Value.

⇓

Variance

$P_0', P_1', P_2', P_3' \; --- \; P_n'$

$x_0, x_1, x_2, x_3 \; --- \; x_n$

② Max Variance $= \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n}$ => Cost function.

Goal : { Find the best unit vector which capture maximum variance }.

(Q) -> How to find the vectors ?

Eigen value Decomposition -> Eigen values and Eigen vectors

$f_1 \quad f_2 \quad$ O/P $\qquad$ 2D → 1D

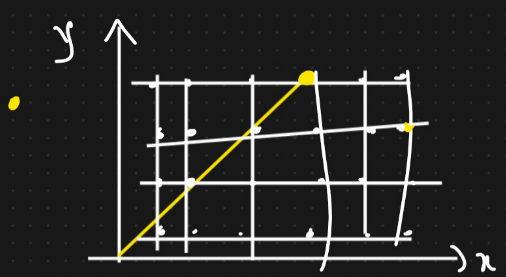
① Covariance Matrix between features $Cov[f_1, f_2]$.

② Eigen value and Eigen will be found out using this Covariance Matrix $\boxed{A v = \lambda v}$

λ → Eigen value

③ Eigen Vector → Eigen Value → Capturing the maximum variance.

Eigen vectors And Eigen Values [Linear Transformation]



$Cov[x, y]$

$\begin{bmatrix} - & - \\ - & - \end{bmatrix}$

$$\begin{bmatrix} & \\ & \end{bmatrix} * [v] = \lambda * v$$

$\lambda \downarrow$ Eigen Value

$\Downarrow$

Magnitude

Eigen vector $\longrightarrow$ Max Magnitude

$\Downarrow$

Max Eigen Value

$\Downarrow$

PC1

Steps to Calculate Eigen Value and Eigen Vector [2d → 1d]

① Covariance of features.

$X, \underbrace{Y,}$ $Z \rightarrow$ o/p

$\Downarrow$

$$Cov(X,Y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{N-1}$$

$$Cov(X,X) = Var(X)$$

$$A = \begin{array}{c|c|c} & X & Y \\ \hline X & Var(X) & Cov(X,Y) \\ \hline Y & Cov(Y,X) & Var(Y) \end{array}$$

$$\boxed{A \cdot v = \lambda \cdot v}$$

3d - 2d

$\Downarrow$

$\lambda_1 \quad \lambda_2$

$\Downarrow \quad \Downarrow$

PC1 PC2.

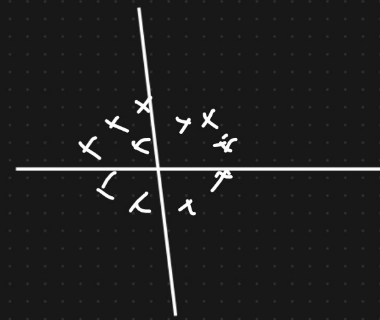
3d - 1d

λ_1

$\Downarrow$

PC1

$\boxed{\lambda_1 \quad \lambda_2} \Rightarrow$ Eigen Values

1  ⇒ Standardize the dataset ⇒ 

2 Cov (x, y)

3 Find out the Eigen value & Eigen vector

$$\boxed{Av = \lambda \cdot v}$$

2 3 . . . n

Variance ↑↑ ⇒ PC1, PC2, PC3 . . .