Principal Component Analysis (PCA)

[Dimensionality Reduction]

① Curse of Dimensionality :

Dimensions → Dataset : 500 features

3 feature → $\boxed{M_1}$ Acc ↑

6 fea → $\boxed{M_2}$ > Acc ↑↑

15 fea → $\boxed{M_3}$ > Acc ↑↑↑

$\boxed{50\ \text{feat}}$ → $\boxed{M_4}$ < Acc ↓↓ → Mode is overfeded

Price of the House

House size

No. of bedrooms

No. of bathrooms

100 fea → $\boxed{M_5}$ < Acc5 ↓↓↓

< 500 featu → $\boxed{M_6}$ Acc6 ↓↓↓↓

* no. of features increasing Model performance Degrade
* features are less but accuracy is high
* feature are more but accuracy is less so mode is overfeded

that's way we reduce the dimention

Domain

$\boxed{\text{Person}}$ ↓↓↓

⇓

Confused — Accurate ↓↓

⇓

not tell price

f_1

2 bhk $\boxed{\text{LOCA}}$ — 450k – 500k

$\boxed{\text{3 bhk}}$ — 500k – 600k

$\boxed{\text{Beach}}$ ← ↑↑

$\boxed{\text{Near to beach}}$ ← ↑↑

$\boxed{\text{Near Grosary shop}}$ ← ↑↑

$\boxed{\text{School}}$ —

Two different ways ways to remove Curse of dimentionality

① Feature selection

⇑

Imp features

② Dimentionality Reduction (PCA)

⇓

Feature Extraction

$f_1\ f_2\ f_2$ o/p

lessor ⇓ ⇓ feature extraction

$b_1\ b_2$ o/p

Feature Selection Vs Feature Extraction

↳ Dimensionality Reduction

① why Dimensionality Reduction ?

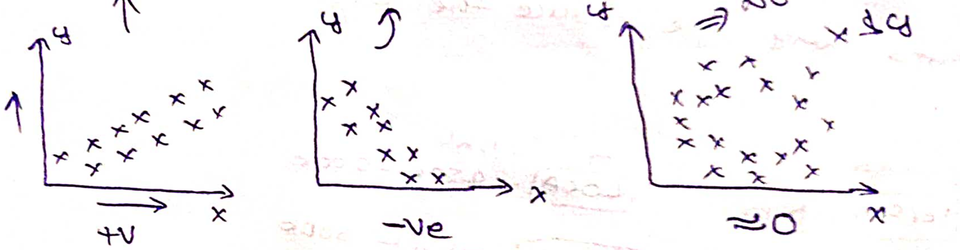
* prevent — Curse of Dimensionality
* improve the performance of the model
* If high features are there then model take more time so we reduce dimention to improve model performance
* Visualize the data → understand the data ⇒ $\boxed{1000}$ ⇓ $\boxed{1d}$ or $\boxed{2d}$

Feature Selection

I/P x O/P y

+V ⇒ $\boxed{x\uparrow y\uparrow \quad x\downarrow y\downarrow}$ $\boxed{x\uparrow y\downarrow \quad x\downarrow y\uparrow}$ ⇐ −Ve

⇒ No relationship x & y



$$\mathrm{cov}(x,y) = \frac{\sum_{i=1}^{n} (x_i-\bar{x}) * (y_i-\bar{y})}{N-1}$$

= +ve ⇒ are super important feature

or

−ve

or

≈0 ⇒ No relationship b/w x & y

$$\text{Person correlation} = \frac{\mathrm{cov}(x,y)}{\sigma_x \, \sigma_y} = \boxed{-1 \text{ to } 1}$$

* the more toward the value of +1 the more +ve correlated x & y is
* −ve correlated
* ≈0 { No Relationship }

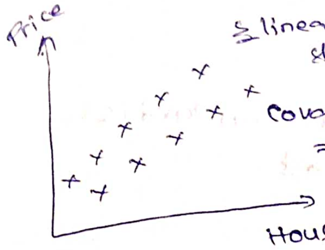
Dataset Housing

→ Independent feature   → o/p feature

House size	Fountain size	Price
—	—	—

Fountain size ↓ Drop

{ feature selection }

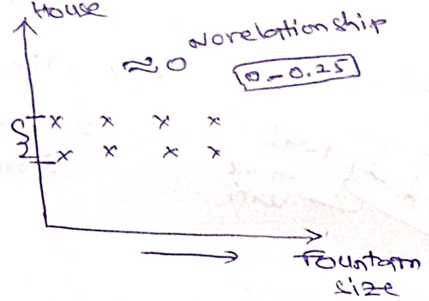


{linear relationship}

Covariance

= +ve

-ve



≈ 0 No relationship

$\boxed{0 - 0.25}$

Feature Extraction :-

Domain Expert

+ve

-ve $\boxed{\text{House size}}$

$\boxed{\text{Price}}$

Room size	No. of Rooms	Price
—	—	—

2 feature → 1 feature

Dimentionality Reduction

⇓ Transformation to extract New feature

House size	Price

PCA Geometric Intuition {Dimentionality Reduction}

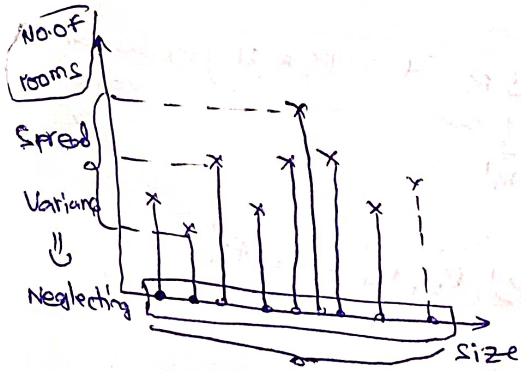
Housing Dataset

Size of House	No. of Rooms	Price

o/p

PCA

2 dimention → 1 dimention



Spread, variance ↑

2D → 1D

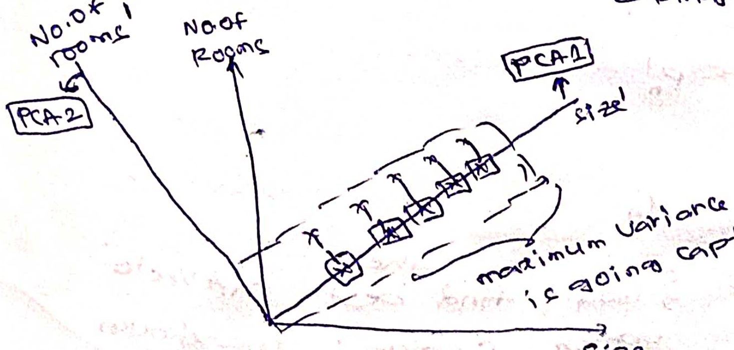
* loss of information (No. of rooms)

⇓

Feature Extraction

* we are taking only House size in 1d
we not selecting No. of rooms

* So we select both 2d into 1d we change the direction axis



$\boxed{\text{Eigen decomposition in matrix}}$

⇑

Transformation

2d ⇒ Pca1, Pca2

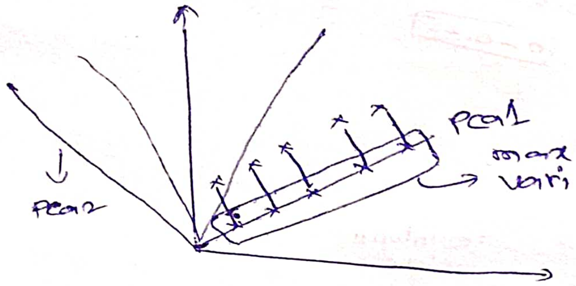
3d ⇒ Pca1, Pca2, Pca3

Variance

* Pca1 maxi Variance captures as Pca2 maxi Vari captures

2d ⇒ Pca1 > Pca2

3d ⇒ Pca1 > Pca2 > Pca3



2D → 1D

2 Best PCA
{ max vari captured is best PCA }

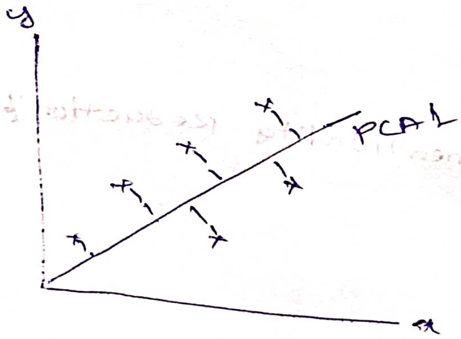
To get the best principal component which captures maximum variance

3D → 1d convert → 1D → * take one best PC1 to 1d

| PC1 |, | PC2 |, PC3 → 2D → take best PC1, PC2 to convert 2d

Var(PC1) > Var(PC2) > Var(PC3)

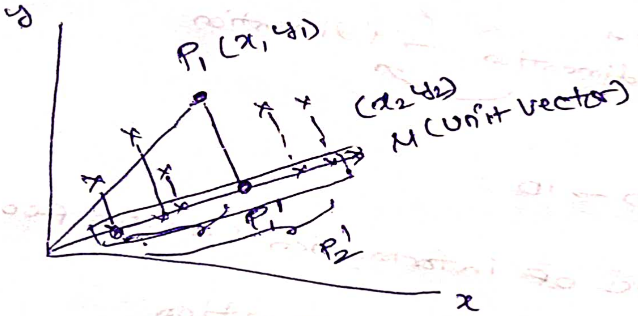
Maths Intuition behind PCA Algorithm



2D → 1D

best PCA or not ?

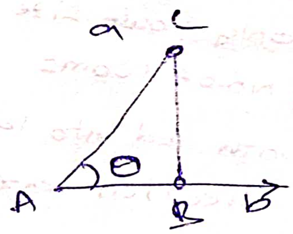
1. Projections
2. Cost function → Variance



$$\boxed{\text{Proj}_{P_1} u = \frac{P_1 * u}{\| u \|}} \quad \|u\| = 1 \text{ unit vector}$$

$$\boxed{\text{Proj}_{P_1} u = P_1 * u} \Rightarrow \text{Scalar value}$$

P_1'



$P_0', P_1', P_2', P_3', P_4' \dots P_n'$

$P_0'\ P_1'\ P_2'\ P_3'\ P_4' \dots P_n'$

⇓

$x_0', x_1', x_2', x_3', x_4' \dots x_n'$

$$\text{Variance} = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n}$$

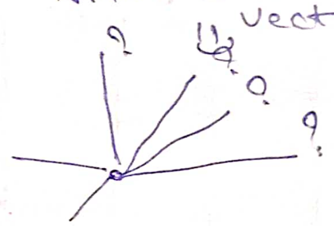
⇓ ... function

{ Goal: find the best unit vector which captures maximum variance }

⇓ using { Eigen Vector, Eigen value }

Eigen Vectors And Eigen Values :-

Find best unit vector



① Covariance matrix between features

② Eigen vectors and Eigen values will found out from this covariance matrix

$\boxed{Av = \lambda v}$

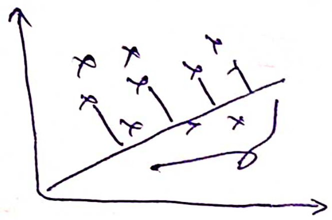
③ Eigen Vector → Eigen value → magnitude of the Eigen vector → capture the maximum varia

Eigen Vectors And Eigen Values [Linear Transformation]

[Eigen decomposition of covari-ance matrix]

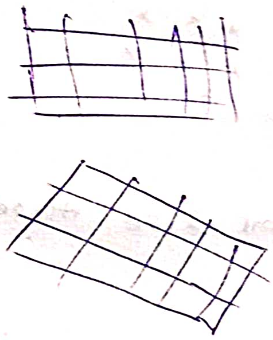
⇓

Eigen Vector & Eigen value



$$\underset{\text{matrix}}{\begin{bmatrix} \; \end{bmatrix}} * \underset{\text{vector}}{[v]} \overset{\text{get}}{=} \lambda * \overset{\text{vector}}{v}$$

λ ⇓ Eigen value



* Apply linear transformation

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

⇓

Eigen vector → maximum magnitude

⇓

Principle compon

⇓

max var

Eigen vector → max magnitude

⇓

max Eigenvector

⇓

Best principle component → PCA

Steps to calculate Eigen value and Eigen Vectors :-

① Covariance of features

$\boxed{x, y}$ → x'

2x2 matrix

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

	X	Y
X	Var(x)	Cov(x,y)
Y	Cov(y,x)	Var(y)

cov(x,x) = Var(x)

Cov(y,y) = Var(y)

	x	y	z
x			
y			
z			

3x3

f_1 & f_2 → λ_1, λ_2 → PC1, PC2

$f_1\ f_2\ f_3$ → $\lambda_1\ \lambda_2\ \lambda_3$

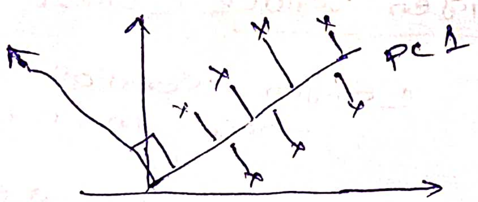
$A \cdot V = \lambda \cdot V$

$\boxed{\lambda_1, \lambda_2}$ ⇒ Eigenvalue

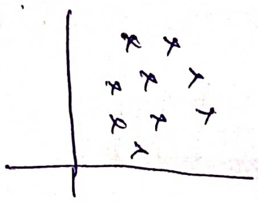
highest select λ_1 or λ_2

⇓

$\boxed{PC1, PC2}$

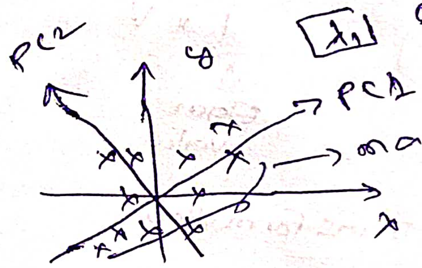


(*)



2D → 1D

① standardize the data



$\boxed{\lambda_1}$ magnitude of Eigen Vector

→ max vari

② Covariance matrix of x & y

$$A = \begin{array}{c|cc} & x & y \\ \hline x & Var(x) & Cov(x,y) \\ y & Cov(y,x) & Var(y) \end{array}$$

2x2 matrix

③ Find out Eigen Vector & Eigen Value

$AV = \lambda V$

$\boxed{\lambda_1, \lambda_2}$ ⇒ Eigen values

↓ PC1 ↓ PC2

3D $\rightarrow$ 1D

λ_1 λ_2 λ_3

$\downarrow$ $\downarrow$ $\downarrow$

PC1 PC2 PC3

1D 1d $= 2D$

3d — 1d

λ_1 λ_2 λ_3

1d

2d $\rightarrow$ 1d

λ_1 λ_2

$\downarrow$ $\downarrow$

PC1 PC2

$\boxed{1d}$ $\Rightarrow$ projection