• Principal component Analysis (PCA) [dimensionality reduction]

① curse of dimensionality.

3 features	4 fea	10 fea	50 fea	100 fea	500 feat
m1	m2	m3	m4	m5	m6

Acc 1 > Acc 2↑ > Acc 3↑↑ < Acc 4↓↓ < Acc 5↓↓ < Acc 6↓↓↓

* if you are feeding model more features (or) Dimensions more then the model will perform bad n t knowing which is more important.

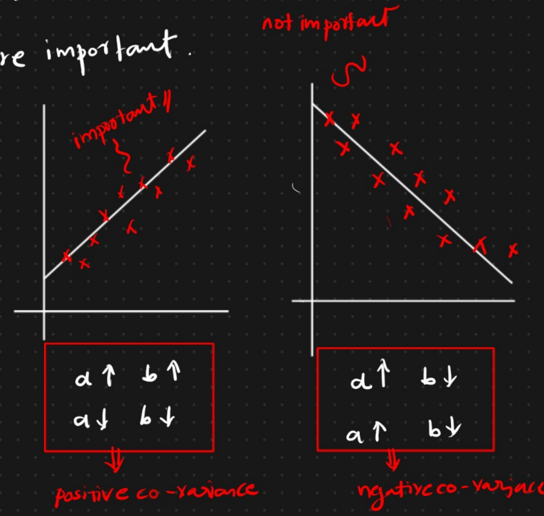
• two diff ways to remove the curse

① feature selection ② Dimension reduction (PCA)

⇑ Imp features ⇑ feature extraction

selecting imp features

* co-variance
* co-relation −1 to +1



② Feature Selection vs Feature Extraction

why dimensionality reduction ??

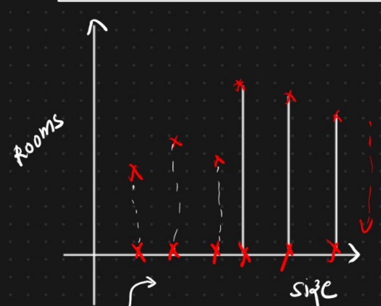
* prevent curse
* improve the Acc ↑↑
* visualize the data

3d 2d ← reduce — 100d wtf ??

we can see

③ PCA geometric Intuition

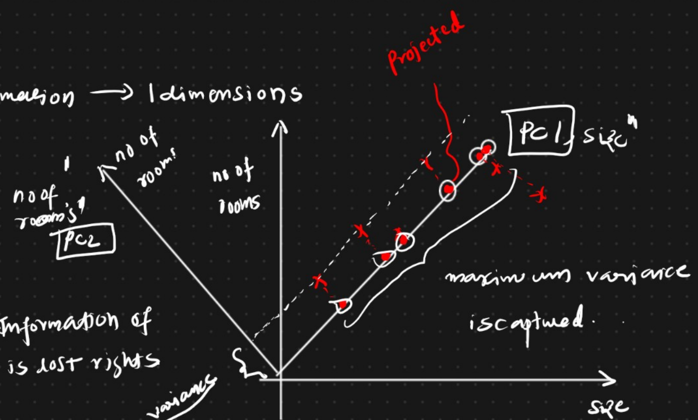
Size of house	No of Room	Price



here Information of Rooms is lost rights

PCA

2 dimension → 1 dimensions



* So, here after projecting the Datapoints the data is getting lost

* So, using Eigen decomposition we create pc1 after projecting it will loose info so,

Eigen decomposition on matrix.

⇑

transform

PC1 → captures the maximum variance.

PC2 → captures the second maximum variance.

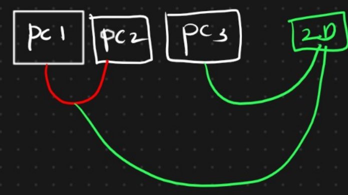
* we create pc2 adcent to the pc1 and this will cover second maximum variance

- 2 Dimensions => pc1, pc2, var(pc1) > var(pc2)
- 3 Dimensions => var(pc1) > var(pc2) > var(pc3)

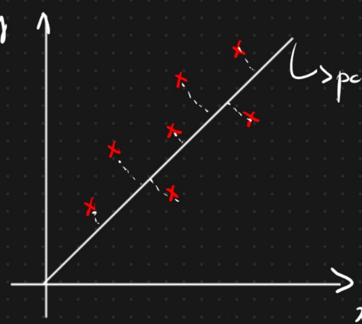
- 2D => 1D

the best pc line draw where is captures more variance (more info)

3D => 2D

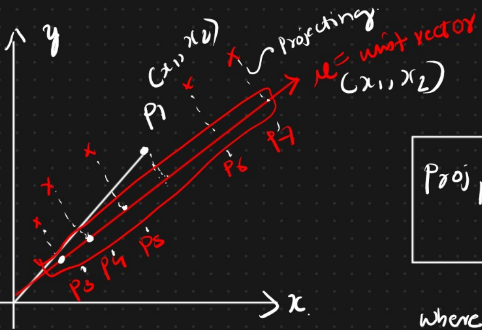


Maths Intuition behind pca Algorithm



to calculate pc it uses:

1. projection
2. cost function (variance based)



$$\boxed{Proj_{p_1} u = \frac{p_1 \cdot u}{\|u\|}}$$

where unit vector magnitude is "1"

$$\boxed{Proj_{p_1} u = p_1 \cdot u} \quad \|u\| = 1$$

$\Downarrow$ p1

scalar values //

variance.

$$\boxed{p_0', p_1', p_2', p_3', p_4'}$$

$\Downarrow$

Scalar values

- (tell's us the only distance)
- this makes it easier to calculate the variance

$$\boxed{x_0', x_1', x_2', x_3', x_4'}$$

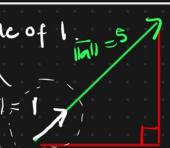
cost function

$$\boxed{Variance = \frac{(x_i - \bar{x})}{N}}$$

here our main goal is to find a best unit vector which captures maximum variance

about unit vector = magnitude of 1

$\hat{u}$ denoted by

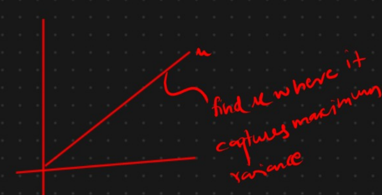


let say we have $\vec{a} = (3, 4)$

magnitude $\|\vec{a}\| = \sqrt{3^2 + 4^2} = 5$ //

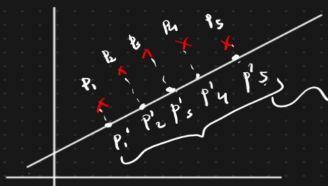
to convert vector to unit vector we divide vectors with their magnitude.

$$\sqrt{\frac{3^2}{1} + \frac{4^2}{1}} = 1 //$$

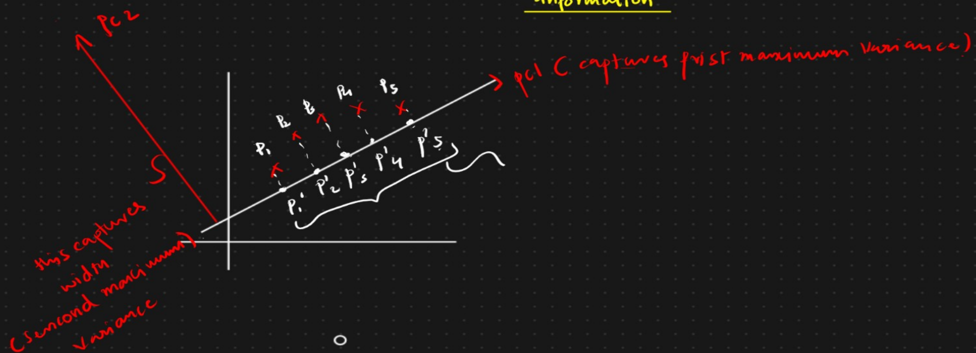


- why do we need pc 2 or pc 3 ?? we have pc3 right ??

let's see



here as you can see we are capturing the only length and we are missing out the width which is nothing but loss of information



Eigen Decomposition on covariance Matrix

= eigen value & eigen vectors

inorder to find the best pc the captures the most variance we use this

- Eigen vector and eigen values [Linear transformation]

[Eigen decomposition of covariance matrix]

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

↓ eigenvalue

$$\begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \text{eigen value} * \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

matrics vector $\hat{i}$

$\boxed{A * v = \lambda * v}$ => eigen vector → maximum magnitude

‖ (mathematically that'll capture maximum variance proven)

Vectors

1) vector multiplication. =

2) Scalar addition. = 2.

vector basics //

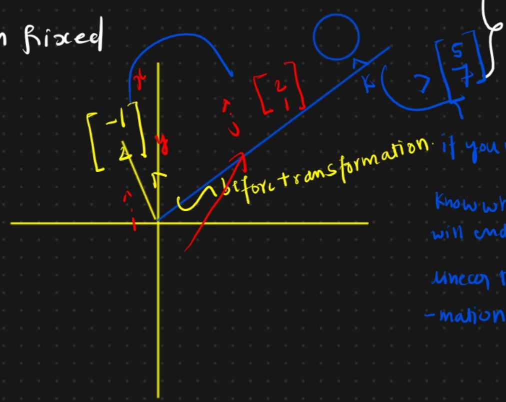
linear transformation

function ⟶ takes vector input and gives an vector output

when the lines after transformation is straight & not bent. and the origin remain fixed

grids are spaced parallel and linearly

numerically ??



if you wanna know where this will end up after linear transformation

$\begin{bmatrix} -1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 5 \\ 7 \end{bmatrix}$

$\hat{i} \quad \hat{j}$

$= 5\begin{bmatrix} -1 \\ 2 \end{bmatrix} + 7\begin{bmatrix} 2 \\ 1 \end{bmatrix}$

$= \left(x(\hat{i}) + y(\hat{j}) \right)$

Steps to calculate eigen vector and eigen values

$$\boxed{\mathrm{cov}(x,y) = \sum_{i=1}^{n=0} \frac{(x-\bar{x})(y-\bar{y})}{n-1}}$$

$$A = \begin{array}{c|cc} & x & y \\ \hline x & \mathrm{var}(x) & \mathrm{cov}(x,y) \\ y & \mathrm{cov}(y,x) & \mathrm{var}(y) \end{array}$$

$\mathrm{var}(x) = \frac{(x-\bar{x})^2}{n-1}$

$\mathrm{var}(y) = \frac{(y-\bar{y})^2}{n-1}$

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

$\boxed{\lambda_1, \lambda_2}$ = eigen values

λ_1 = magnitude of pc1

λ_2 = magnitude of pc2.