Welcome back everyone.

⇒ PCA
⇒ Projects → end to end

⇒ DL / NLP

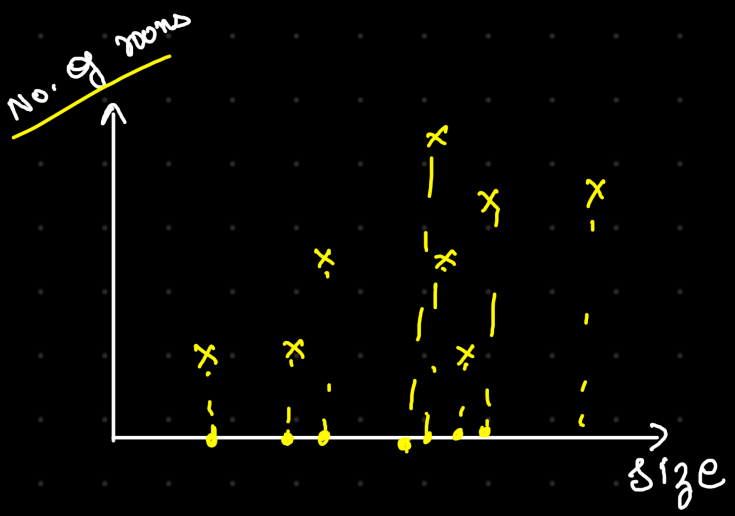
✓ Git
✓ Web Scraping

Principle Component analysis (PCA)

size of rooms | No. of rooms | Price

reduce the dimensionality

2 independent → 1 independent

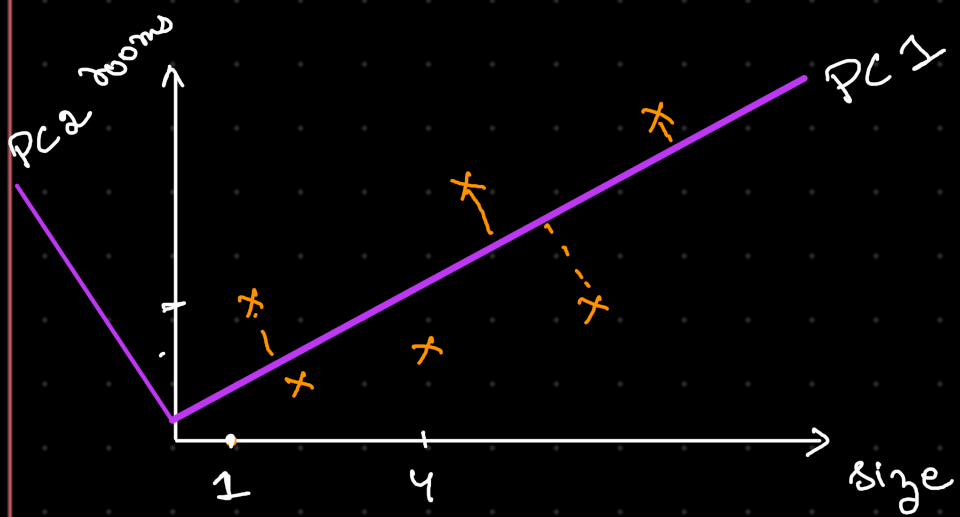


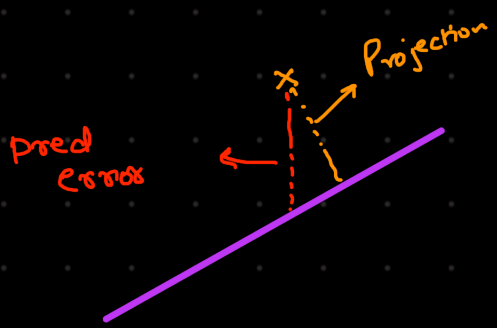
① Feature selection

2D ⇒ 1D

Problem is that all information about no. of rooms is lost

All info lost





Now, will be able to capture a lot more variance.

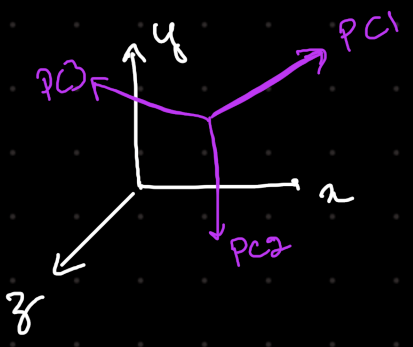
spread and variance of both room & size.

no. of features : no. of Principal Component

variance / spread } information of data

3 dim^n → PC1 PC2 PC3 always ⊥^{r}

will take first 2

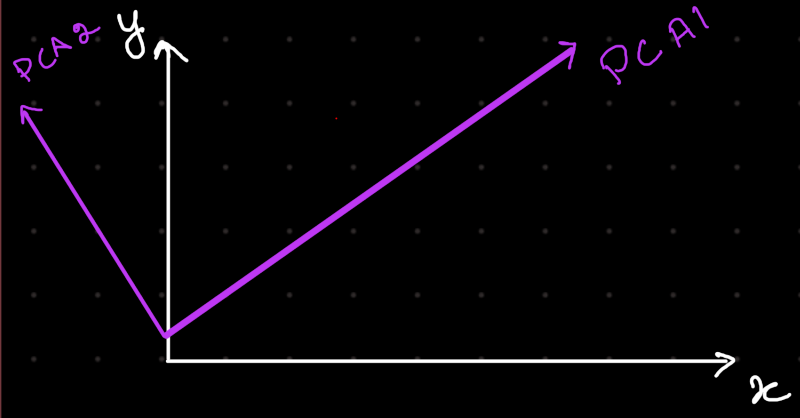


reducing dimensionality of my data

10000dim → 10 dim

rooms → 1 rooms

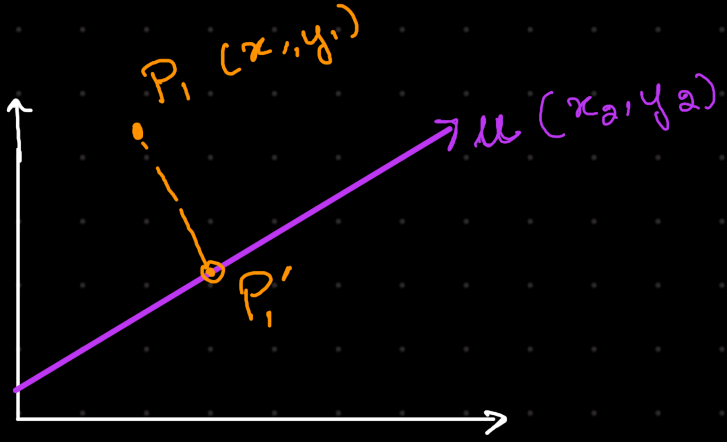
Mathematical Intuition of PCA

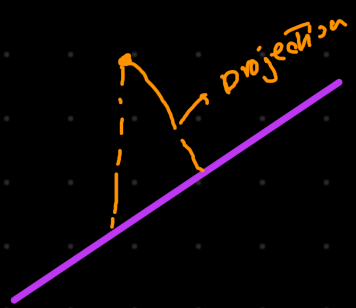


2 features
⇒ 2 PCA vectors

① Projections ? we are projecting our data point

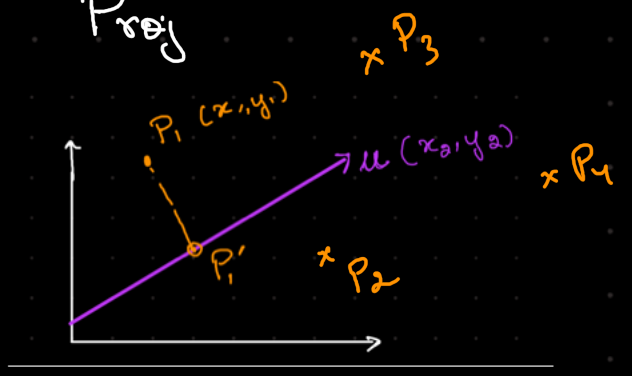
② Optimization ? which vector / PCA gives us max^m variance.





P_1' will be my final point which I will be using in my model training.

Proj



$$\text{Proj}_{P_1} u = \frac{P_1 \cdot u}{\|u\|} \rightarrow 1 \text{ for unit vector}$$

$$= P_1 \cdot u$$

$$= \begin{bmatrix} x_1 \\ y_2 \end{bmatrix} \cdot [x_2, y_2]$$

$$= x_1 x_2 + y_1 y_2 \Rightarrow \text{scaler value.}$$

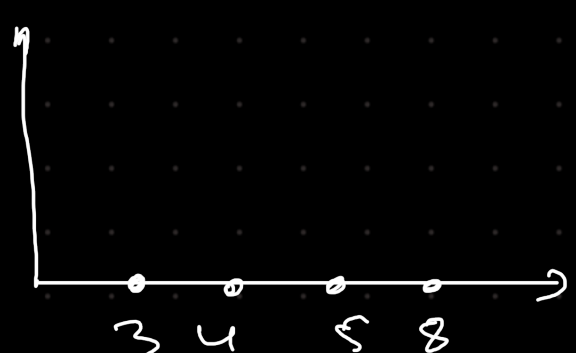
P_1 P_2 P_3 P_4
↓ ↓ ↓ ↓
P_1' P_2' P_3' P_4'

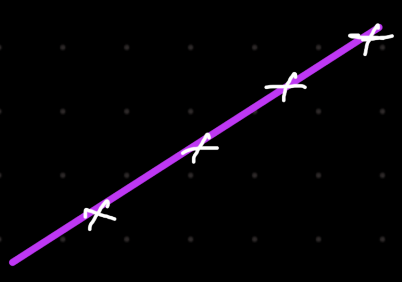
⇒ scaler value and can then calculate the variance

x_1' x_2' x_3' x_4'

(2) Maxm variance

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





How do we find u ?

Eigen values decomposition

↙ eigen value

↘ eigen vector

Eigen values tells how much a specific eigen vector is capturing the variance

high eigen value → high variance factor.

Steps.

(1) Covariance matrix b/w features

$$Cov\{f_1, f_2\} \Rightarrow A$$

(2) Eigen value & eigen vector will be found out using covariance matrix

$$Av = \lambda v$$

A ↙ Cov matrix

λ ↓ eigen value.

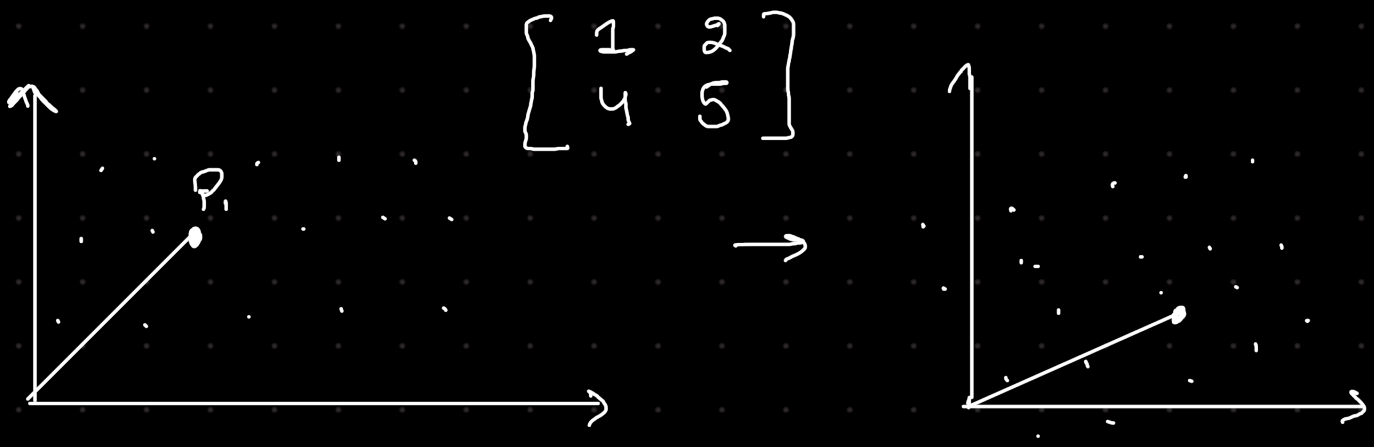
v → vector

(3) Eigen vector with most eigen value will be selected as that will capture highest variance.

How to calculate eigen vector & eigen value

[Linear Transformation] of vector

$$\begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix}$$



eigen vector is one where our dirn remains same

we have to find

eigen value ⇒ value of vector

$$\begin{bmatrix} & \\ & \end{bmatrix} \cdot v = \lambda v$$

Cov Matrix

λ ↓ eigen value

eigen vector → max magnitude

⇓

max eigen value

⇓

PC1

Steps to Calculate eigen value and eigen vector

2d → 1d

① Always standardize before applying PCA

② Covariance of features

x_1 x_2 t_1

⇓

$$Cov(x_1, x_2) = \sum_{i=1}^{n} \frac{(x_i - \bar{x})(y_i - \bar{y})}{n}$$

Cov

	x_1	x_2
x_1	$Var(x_1)$	$Cov(x_1, x_2)$
x_2	$Cov(x_2, x_1)$	$Var(x_2)$

⇒ A

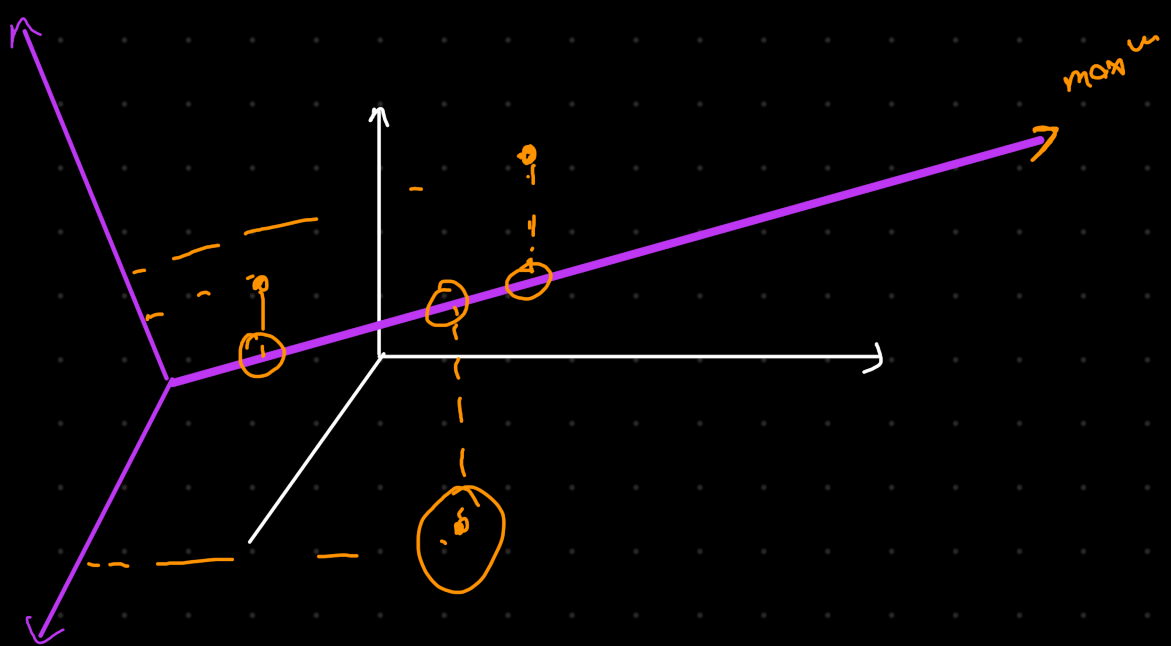
$$A \cdot v = \lambda \cdot v$$

3d → 2d

λ_1 λ_2

⇓ ⇓

PC_1 PC_2



Projects

Python
Statistics
Data Science
ML
DB API, file
Git
Web scraping

DL/nlp ⇒ Jumbo

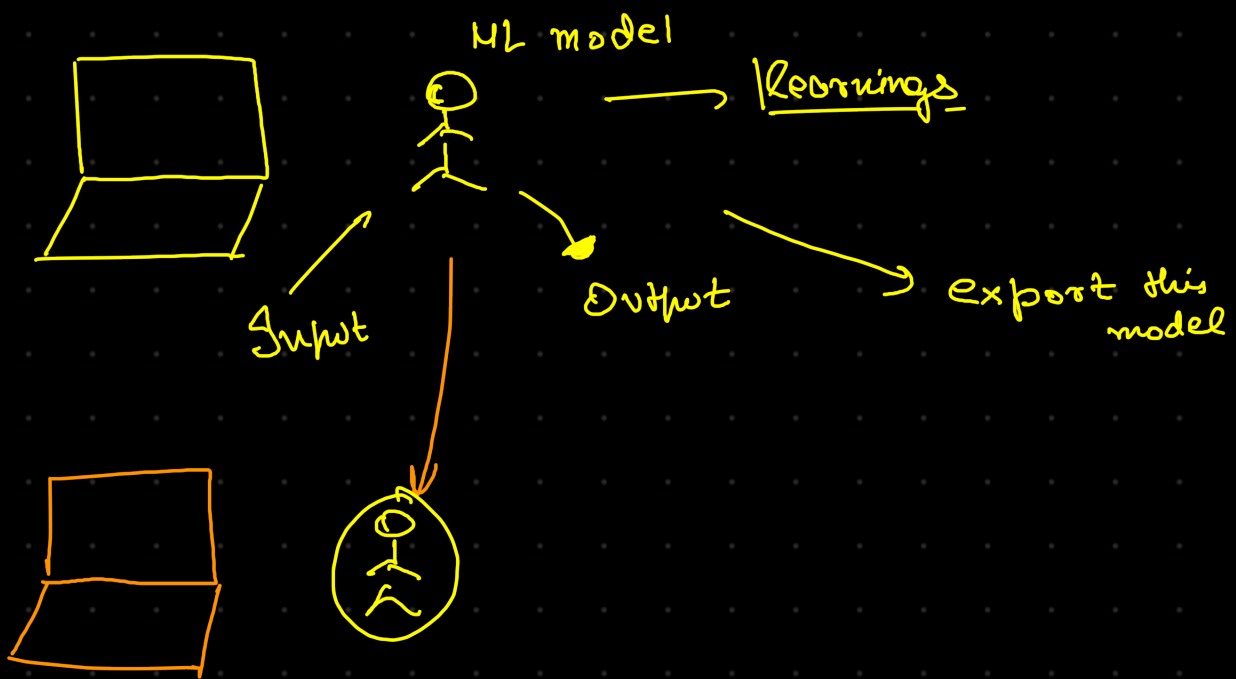
Projects

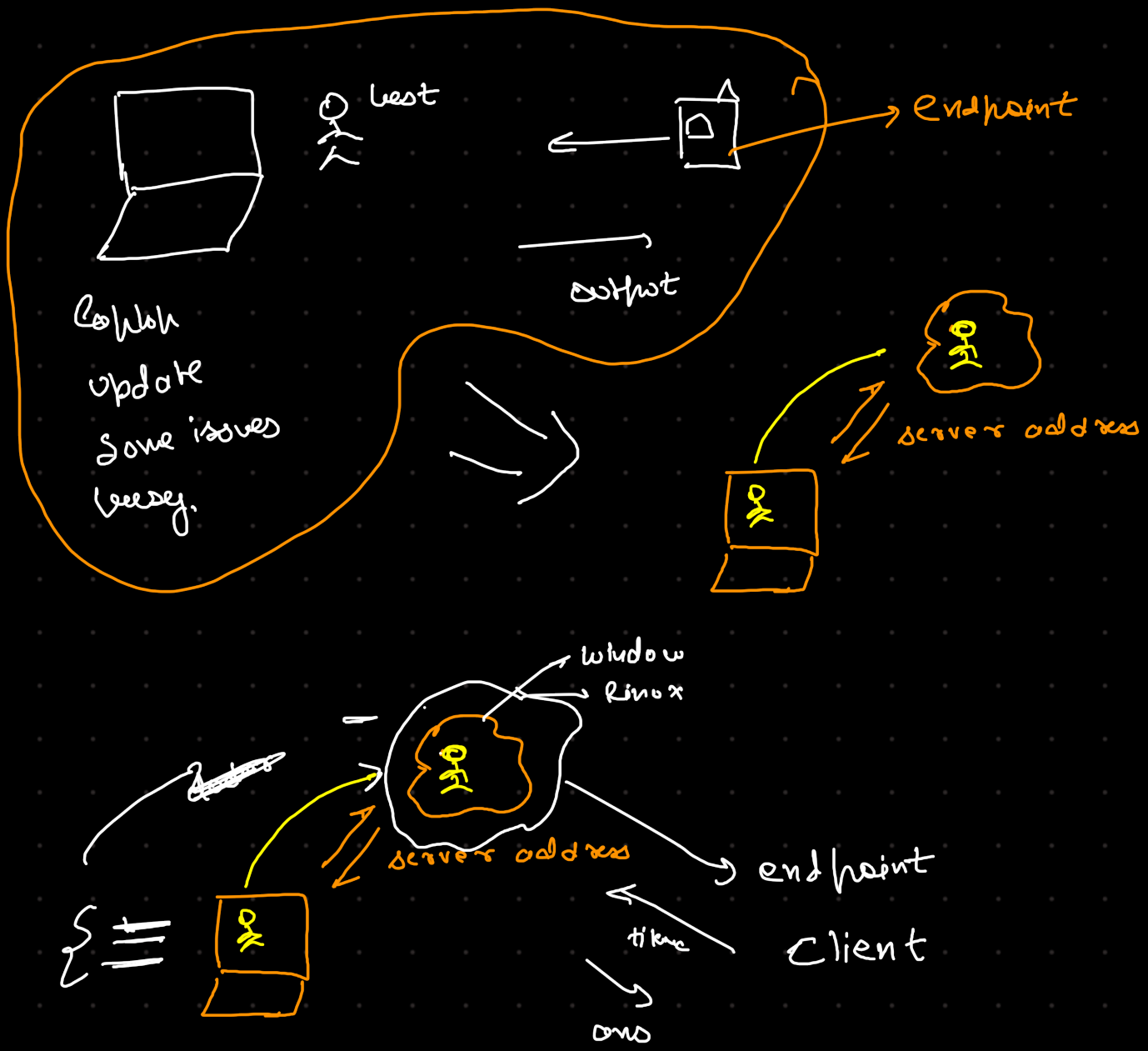
.ipynb

80%

Kaggle
Colab

First we will do projects on .ipynb to have data science concepts clear.




test
endpoint
output
server address
window
linux
server address
endpoint
client
time
one