Principal component Analysis (PCA — A dimensionality reduction techique)

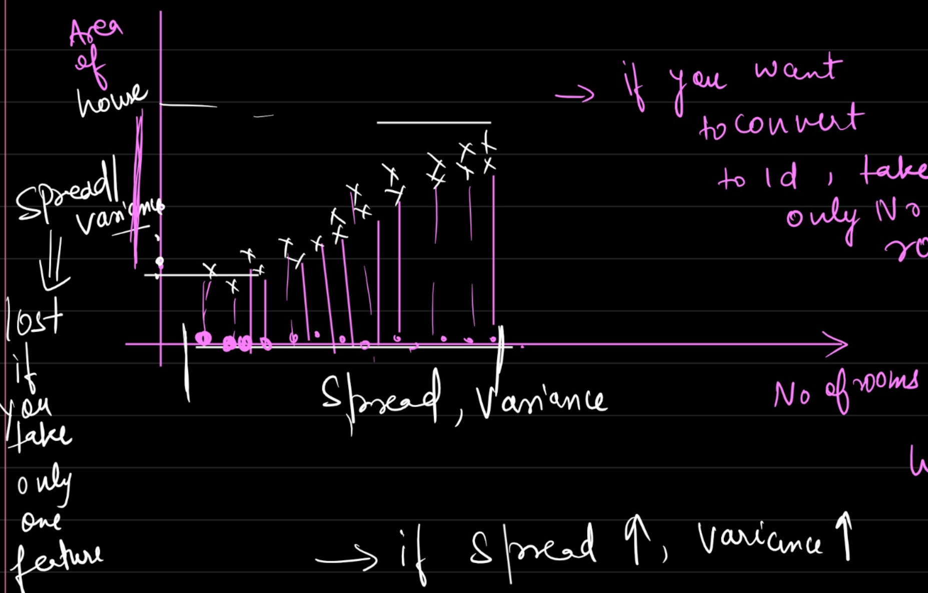
(Geometrical Intuition)

Area of house	No of Rooms	Price of house

* Since PCA is a dimensionality reduction technique → we will try to reduce the dimension

$\boxed{2d \longrightarrow 1d}$ → Using PCA.

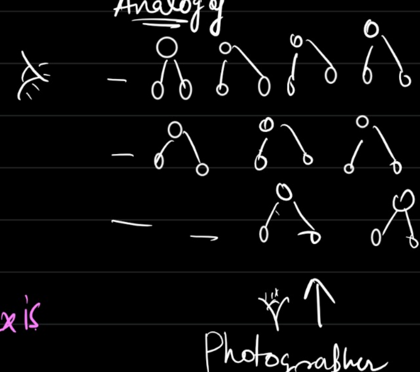
* ML is all about learning patterns of data.

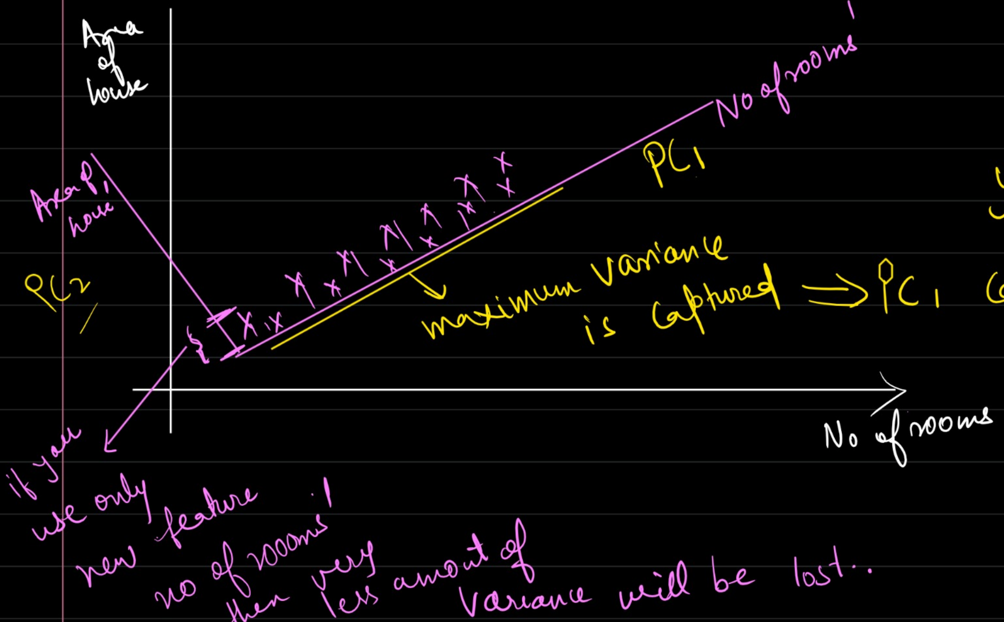


→ if you want to convert to 1d, take only No of rooms
⇓
x-axis
⇓
We will loose area of house (Actually this is happening in feature selection)
⇓
You want both the features.

→ if Spread ↑, Variance ↑

Analogy



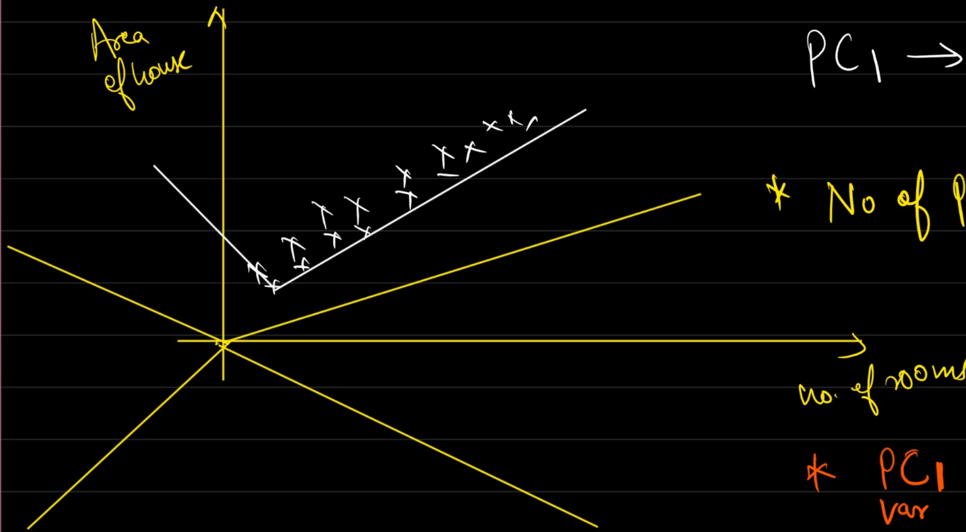


maximum variance is captured ⇒ PC1 can be used for model training.

2d — PC1 PC2

PC2 is used because it captures maximum spread of data

2d — 1d (PC2)



$PC_1 \rightarrow$ maximum Variance (spread)

* No of principal components will be equal to no of features.

* $\underset{Var}{PC_1} > \underset{Var}{PC_2} > \underset{Var}{PC_3} > \underset{Var}{PC_4} \dashrightarrow \underset{\overline{\overline{Var}}}{PC_n}$

f_1 f_2 f_3 ---- f_{1000}	y	PC_1	PC_2	PC_3	---- PC_{1000}

f_1 f_2 f_3

| | |

PC_1 PC_2 PC_3

$var(PC_1) > var(PC_2) > var(PC_3)$

axis transformation $\Rightarrow$ $\boxed{\text{Eigen decomposition of Covariance Matrix}}$