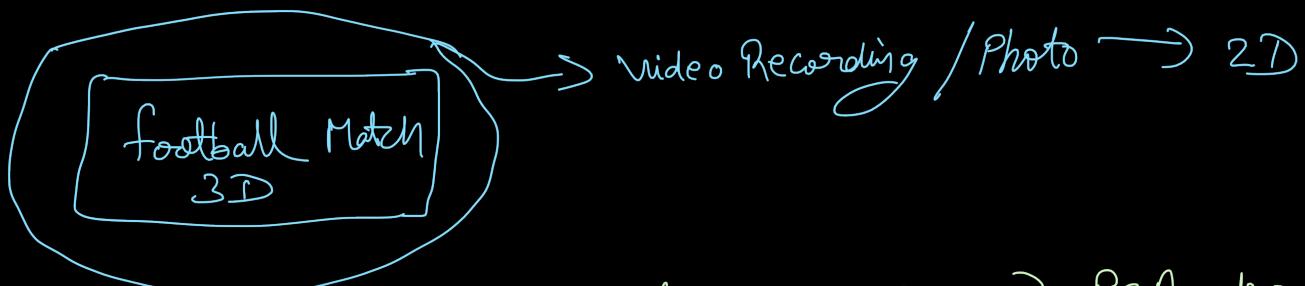


# unsupervised $\hookrightarrow$ Principle Component Analysis (PCA) [Geometric Intuition] Machine Learning Method

## Introduction

feature extraction technique  $\xrightarrow{\text{to reduce}} \text{Curse of dimensionality}$



3D moment to 2D  $\Rightarrow$  PCA also exactly does the same thing.  
Higher dimensional data set  $\rightarrow$  Lower dimensional data set

PCA is a technique which can transform higher dimensional data to lower dimensional data while keeping the essence of the data.

(Creating new set of features from existing ones in such a way that more variance can be captured from the higher dimension)

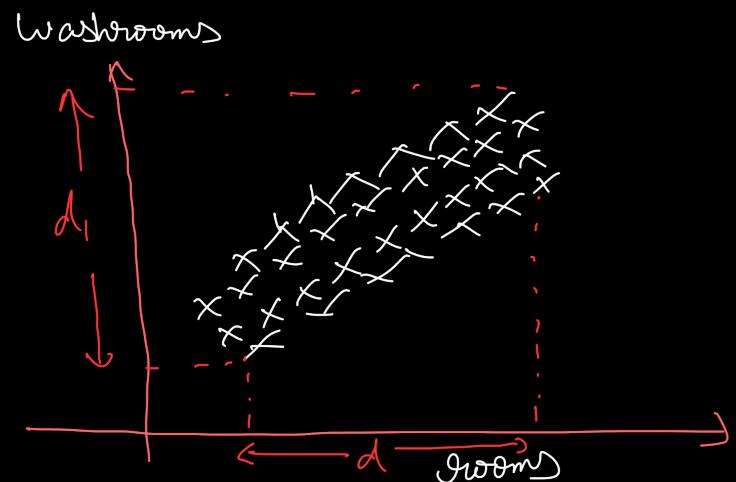
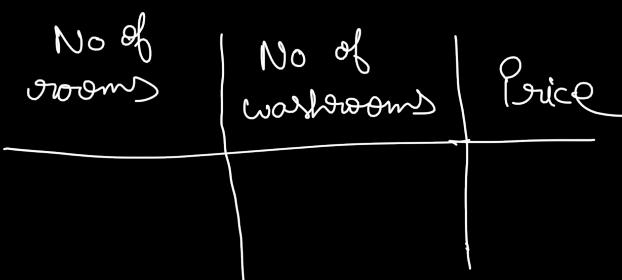
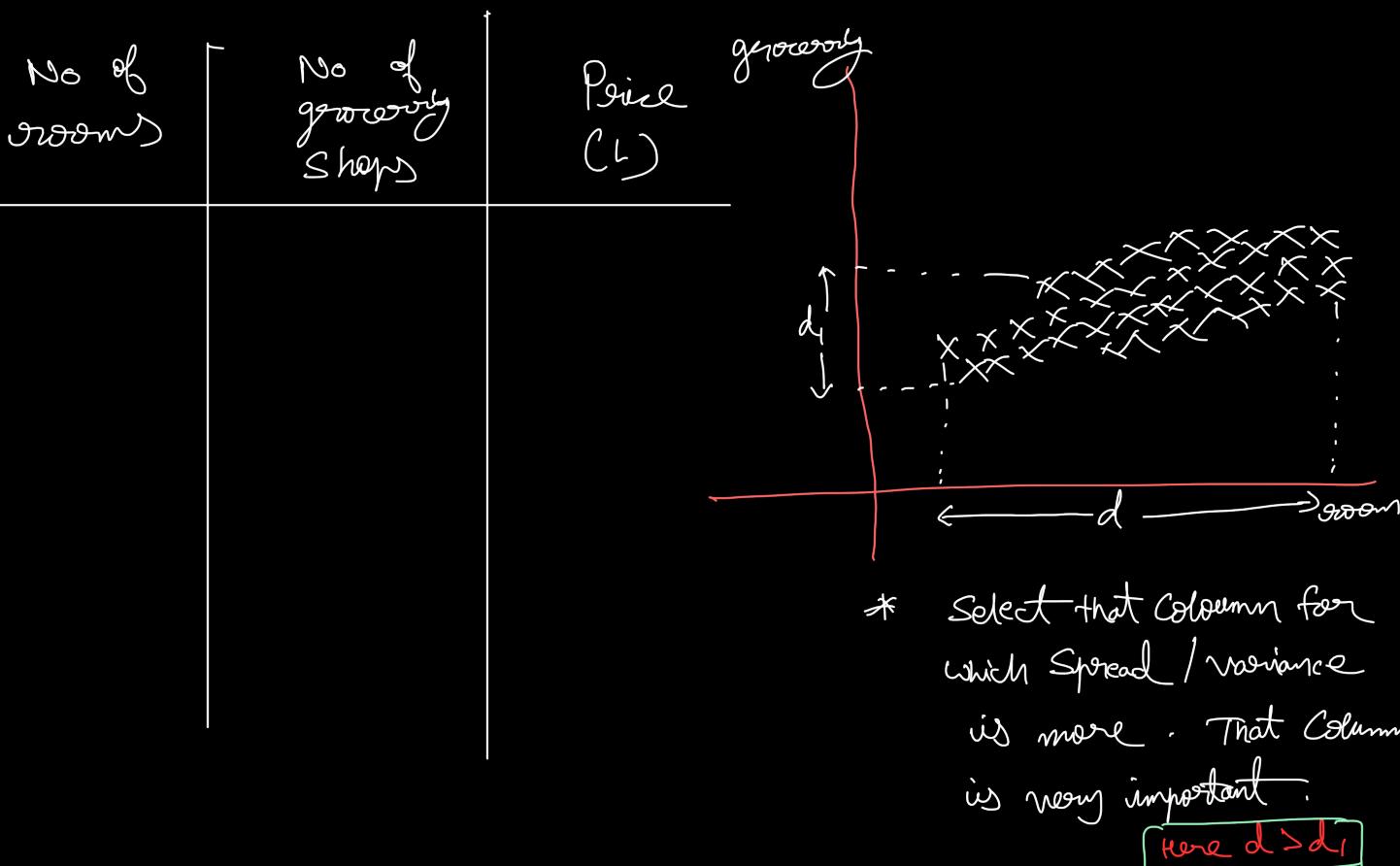
(best possible lower dimension so that essence of data in higher dimensions can be captured)

Core Benefits : 1) Higher to lower (in terms of dimension)  
So faster execution of algo.

2) we can visualize by lowering dimension to 2D or 3D.

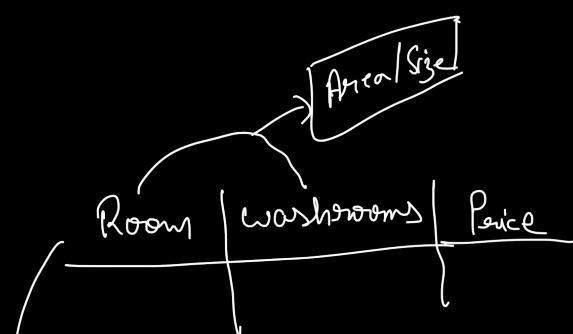
# Geometric Intuition

## feature selection



Both variable/features are linearly related to each other.

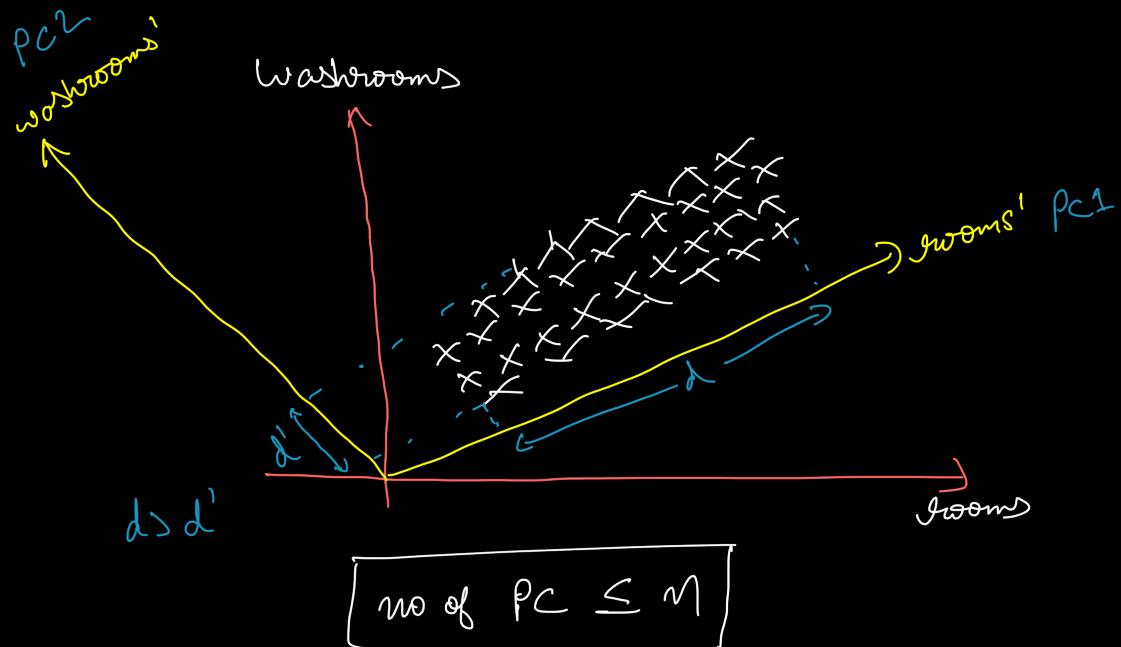
In this case feature selection will not able to help us. Here feature extraction will help.



This is what PCA does. From existing set of features, it creates totally new set of features.

PCA

- 1) will find a new set of axis by rotating existing coordinates
- 2) transform data according to PC1 because  $d > d'$ .



Why variance is Important?

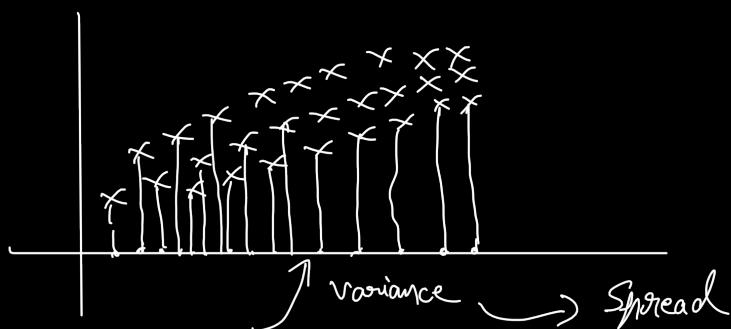
$$\text{Variance} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

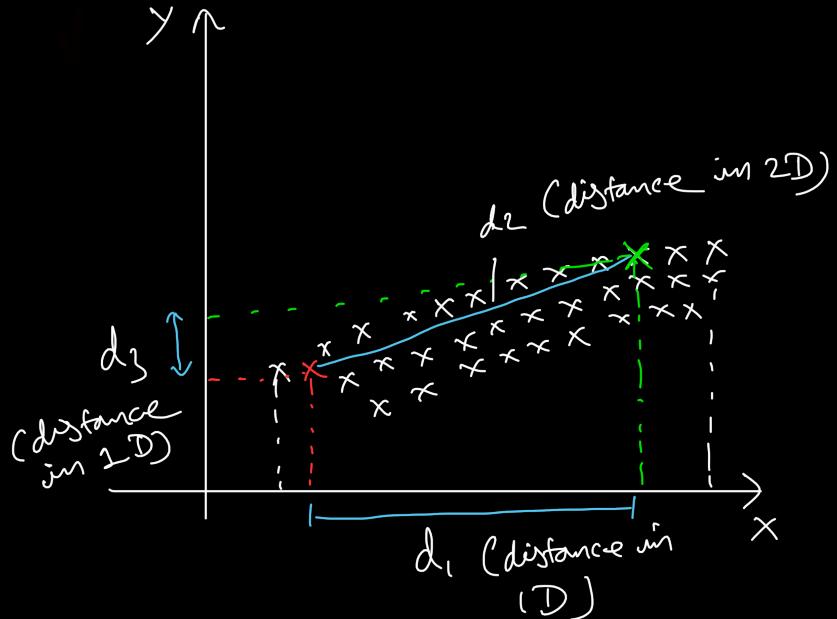
Spread is not variance

Spread & Variance

Mean absolute deviation = Spread, we don't use MAD / Spread (MAD)  
directly in PCA because it is not differentiable at 0 so it will not converge in time of optimization

Mod





(i) when projecting in X-axis then atleast we have comparable distance between red and green point from their 2D representation. see  $d_1$  &  $d_2$  they are almost same.

(ii) But now when we do projection of points on Y-axis then  $d_3$  is smaller than  $d_2$ .

{ That is why variance is important. when we bring data points from higher to lower dimension, Relationship between data points do not get distorted, to maintain this we have to maximize the variance. }

End Goal to maximize the variance