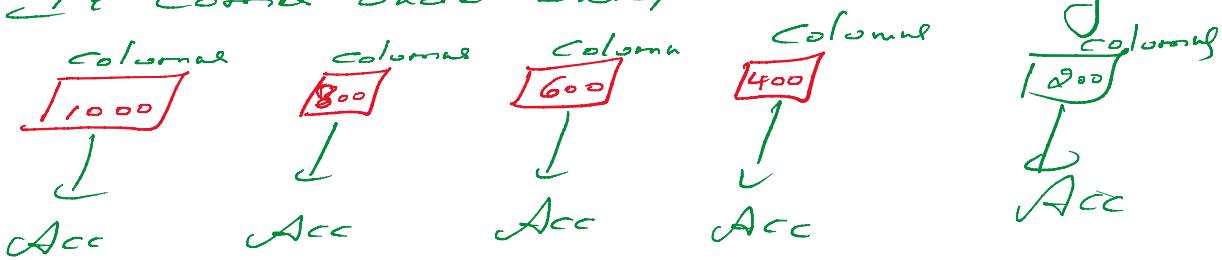


PCA

19 June 2023 21:48

PCA → Principal Component Analysis

* It comes under unsupervised learning.



Curse of dimensionality

* As the number of features increases the model performance decreases

* Computation time is high.

* Hard to visualize

* Results in overfitting

* Model performance will be less.

How to overcome the Curse of dimensionality?

* Feature Selection \rightarrow Heat map → Correlation
Selecting best features.

* Feature Extraction.

\rightarrow Dimensionality reduction.

PCA

PCA transforms higher dimension data into low dimension and creates new co-ordinate system of points which are called as principal components which explains the maximum variance.

Location / Sqft / BHK / No rooms / no of floors / no of colors / no of shops nearby / price

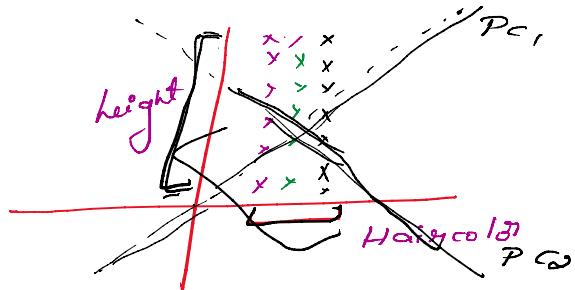
* It drops the columns which has less variance

* It drops constant columns

* It will use eigen vectors, eigen values and covariance matrix to construct new features

1st \rightarrow $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ \rightarrow P_1

Hair / height



Hair/height

$PC_1 | PC_2 | PC_3$

Principal Components
which separate all maximum variance.

$PC_1 | PC_2 | PC_3 | PC_4 | PC_5 | PC_6$

→ Explained variance ratio