***Principal Component Analysis***

*PCA is trying to reduce Dimensions of Data lyk if 10 Dimension dataset hai so ushe 2 Dimension dataset mai leke aana*

*Mtlb ki High Dimension se low dimension mai leke aana while maintaining essence of data.*

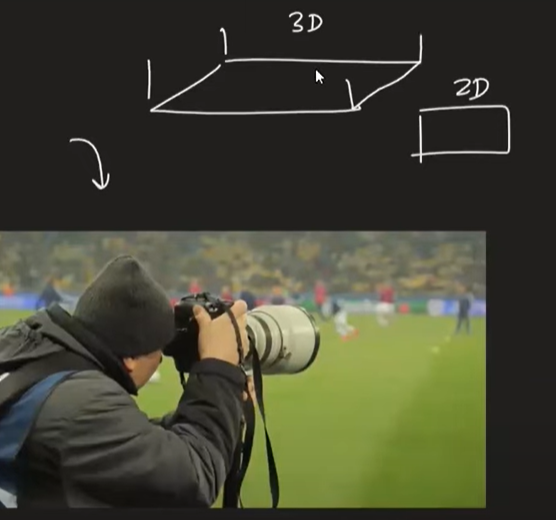
Bht purana technique hai bht reliable hai 1950s mmai kch tou paper aaya tha

Iska Maths thoda complex hai

Feature Extraction Technique hai

Jbh Hum PCA apply krte hai tou

***Analogy***



Yeah Cameraman kya krrha hai ki 3D area mai 2D pictures capture krrha hai.

& yeah photographer move krte rehta hai to take Best Shot of Game.

PCA is a technique which can transform Higher Dimension data into Lower Dimension data by keeping the essence of the data.

Data ka jo Behavior hai vo bachake lower Dimension mai leke aajate hai.

So Lower Dimension ka data acha hie results dega even.

**2 Core Benefits:**

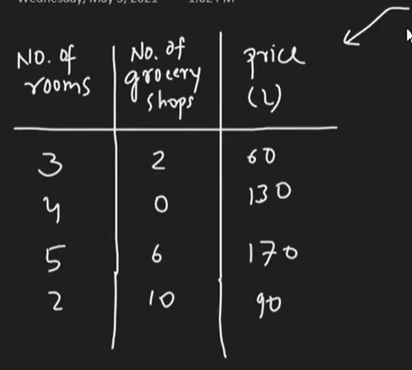
PCA aapke Higher Dimension dataset ko lower Dimension mai lekr aata hai.

Iska size kamm hojaraha hai interms of Dimension

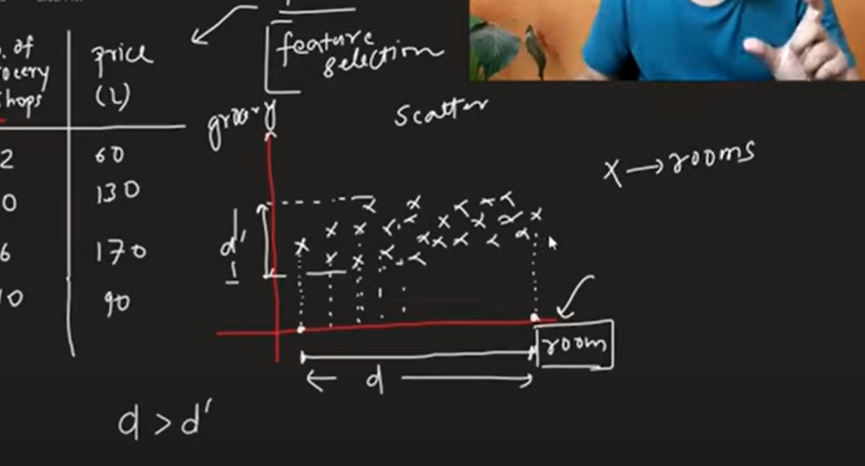
1. Faster Execution
2. Visualization

PCA helps lengthy data ko 2D ya 3d mai reduce krke hum graphs & all plot krskte hai

*Geometric Intuition*



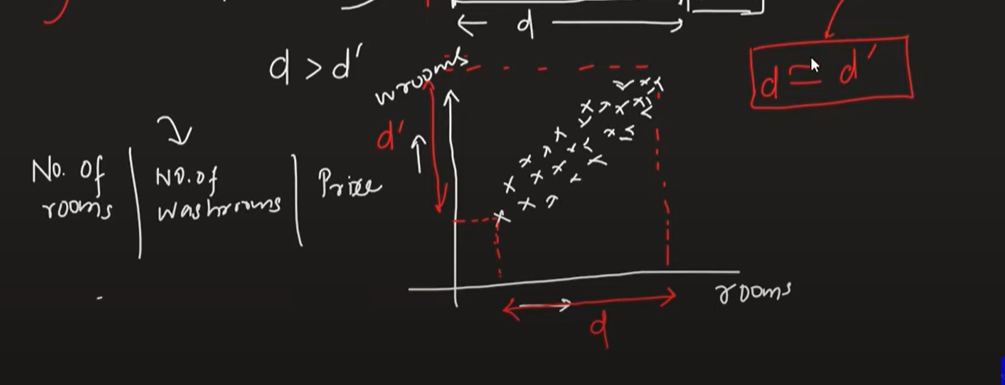
Feature Selection ka scene Ya Problem.



As we can see Rooms jo hai vo zyaada spreadout hai so vo zyaada Important hai

Jis column pe Spread zyaada aata basically variance then vahi Column Important hai sbhse.

**Lets say Data mai changes aagaye**



Abh yaha we can see that No. of Rooms & No. of Washroom dono ke beech Linear Relation jaisa hota hie hai

Jitna bada ghar utna zyaada Washroom

So dono ka spread 1 hie hai dono ka variance basically Equal hai.

So ese jageh pe Feature Selection mai Problem aati hai

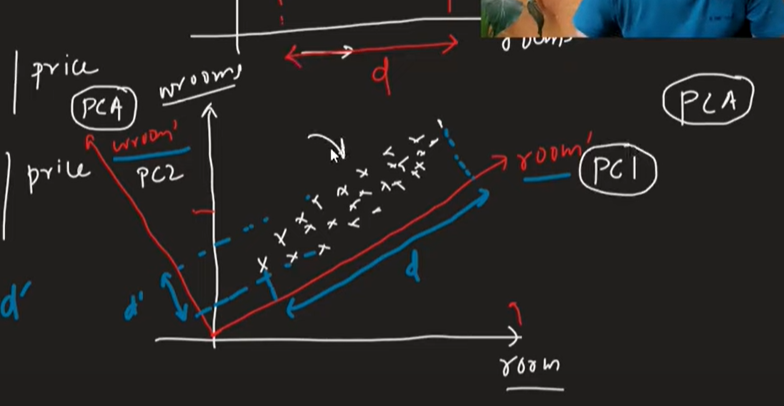
Yahape Humme Help karta hai

**Feature Extraction**



PCA aapke existing Features ko Bhulljaata hai basically

& unnh existing features mai se naye set of features mai se subset choose krta hai Jo ushe lagta hai jo sbhse zyaada Important hai.



Jitne columns honge utne Principal components banenge

Eg upar 2 columns hai tou 2 PC bane hai

No of PC Component is =< than N

Hum ese tarike Se Coordinate ko Transform krte hai ki

1 Particular Component ka variance zyaada aaye.

**Q. Variance Data ka Spread Important kyu hai ??**

Variance is a statistical technique jishse hum Data ka Spread Smjhte hai.

Variance is proportion to spread.

While Principal Component Analysis (PCA) is a powerful technique for dimensionality reduction and feature extraction, there are some situations when it might not be appropriate or might not provide significant benefits. Here are some scenarios when you might want to reconsider using PCA:

1. **\*\*Interpretability:\*\*** PCA creates new components that are linear combinations of the original features. These components might not have a direct interpretation in the original feature space. If interpretability is essential for your analysis or domain, PCA might not be the best choice.

2. **\*\*Non-linear Relationships:\*\*** PCA is effective in capturing linear relationships between variables. However, if your data exhibits non-linear relationships, PCA may not be the most suitable method. In such cases, nonlinear dimensionality reduction techniques like t-distributed Stochastic Neighbor Embedding (t-SNE) or manifold learning methods might be more appropriate.

3. **\*\*Preserving Specific Features:\*\*** PCA aims to retain variance in the data, which means it focuses on preserving the overall structure of the data rather than individual features. If you have specific features that are critical for your analysis, using PCA might lead to information loss for those features.

4. **\*\*Outliers:\*\*** PCA is sensitive to outliers in the data. If your dataset contains outliers, they can significantly influence the principal components and potentially distort the results. In such cases, it might be better to consider robust PCA techniques or outlier removal methods before applying PCA.

5. **\*\*Small Dataset Size:\*\*** In situations where the dataset is small, using PCA may not be very beneficial. PCA relies on the covariance matrix, which may not be well estimated with limited data points, leading to less reliable results.

6. **\*\*Linearly Separable Data for Classification:\*\*** If your dataset is already linearly separable, applying PCA might not provide substantial improvements, as the linear decision boundaries of the original data might be sufficient for classification tasks.

7. **\*\*Time Complexity:\*\*** PCA can be computationally expensive for large datasets, especially if the number of features is high. In such cases, approximate or incremental PCA methods might be considered.

In summary, while PCA is a valuable tool for dimensionality reduction and has many use cases, it is essential to consider its limitations and whether it aligns with the specific characteristics and goals of your dataset and analysis. Depending on your data and objectives, other techniques, such as feature selection methods, manifold learning techniques, or domain-specific knowledge, might be more appropriate choices.