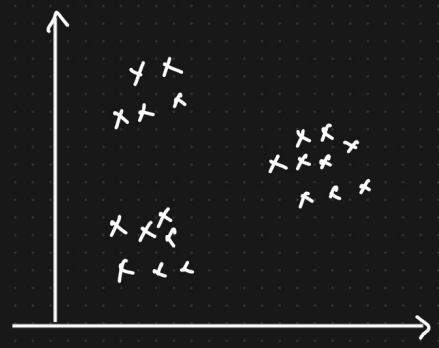
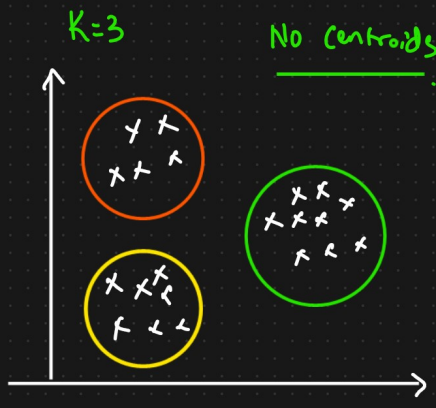
# Hieraridal Clustering



⇒ Hieraridal Clustering

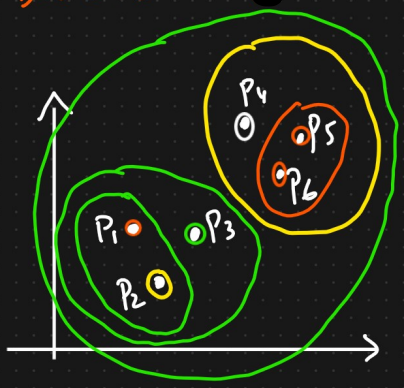



## HC

Combining

① Agglomerative (Bottom-Up)

② Divisive (Top-Down)



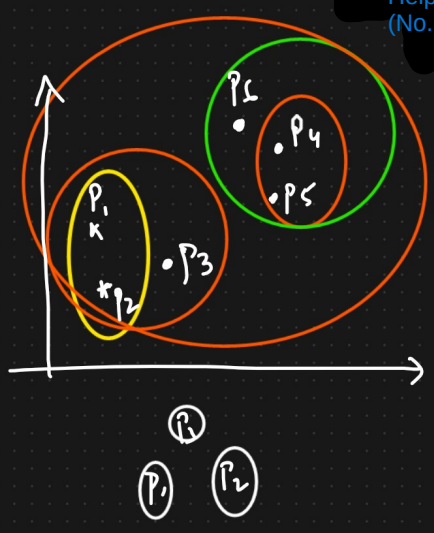

### Steps For Agglomerative :-

① For each point initially will consider it as a separate cluster

② Find the nearest point and create a new cluster

③ Keep on doing the same process until we get a single cluster.

### Dendogram




Helps to find k value (No. of cluster)

Threshold

K=2

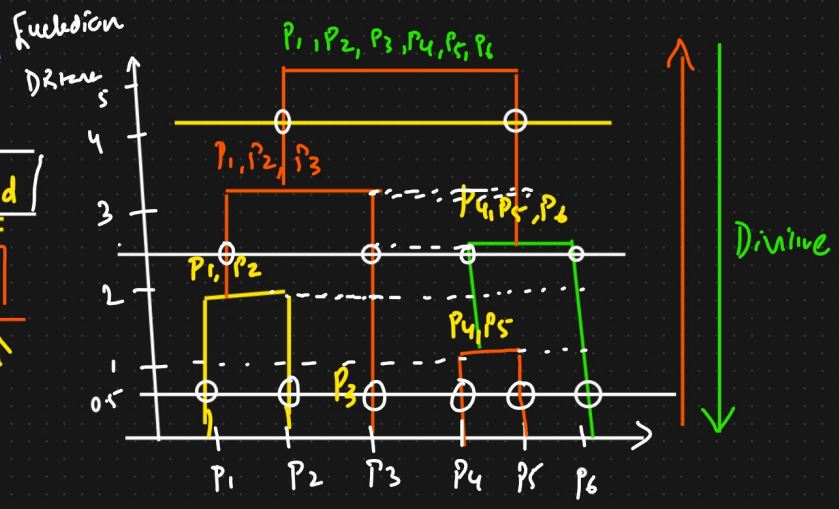
K=4

### Dendogram

K=3

K=2

Agglomerative



(K) Select the longest vertical line such that no horizontal line passes through it → For threshold

Threshold {Euclidean Distance}.

★ Draw a horizontal line that cuts across the largest gap — number of vertical lines it intersects = number of clusters k.

## K Means Vs Hierarchical Clustering

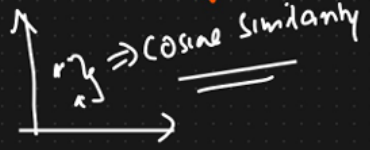
### Scalability And Flexibility

① Dataset size → Huge → K Means

Small → Hierarchical Clustering

② KMean → Numerical data

Hierarchical clustering → Variety of data.




Cosine Similarity :-

- A measure of how similar two vectors are, based on the angle between them — not their magnitude.

$$\cos(\theta) = \frac{A \cdot B}{|A| \cdot |B|}$$

- 1 → Very similar
- 0 → No similarity
- -1 → vectors are opposite.

③ Centroids → Elbow method → No. of centroids

→ No. of Clusters