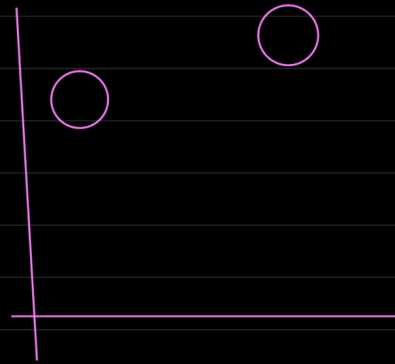
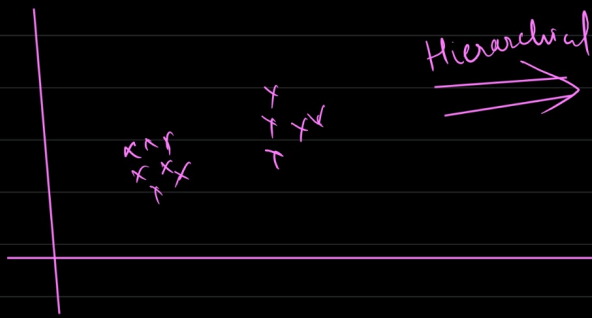


## \* Hierarchical clustering



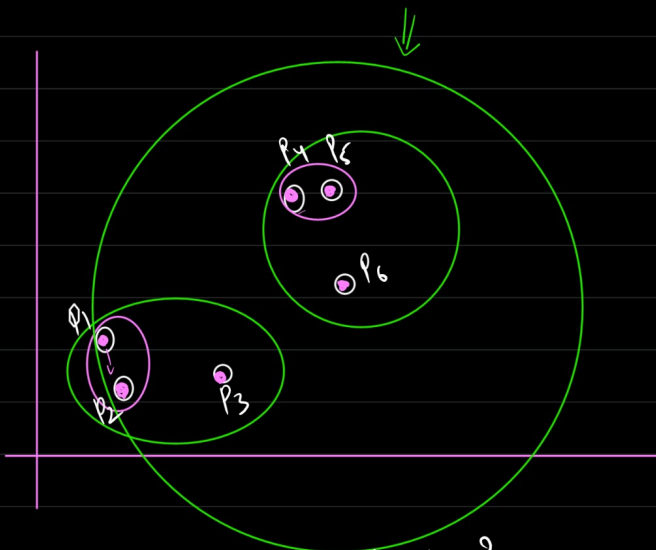
Hierarchical

- ① Agglomerative clustering
- ② Divisive clustering

## \* Agglomerative (Combining)

Steps (Agglomerative)

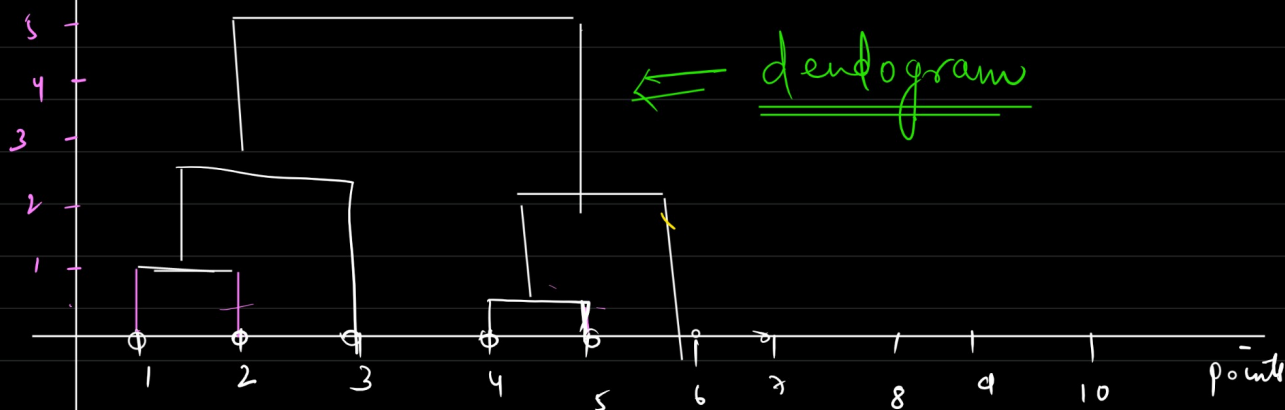
- ① Each point is a cluster in its own.
- ② Find the nearest point and create a new cluster
- ③ Keep on doing step 2 until we get a single cluster.



How you are combining?

- Euclidean dist
- Manhattan distance
- Cosine similarity → To calculate distance b/w two vectors (Categorical / String dist)

Euclidean  
distance



← dendrogram

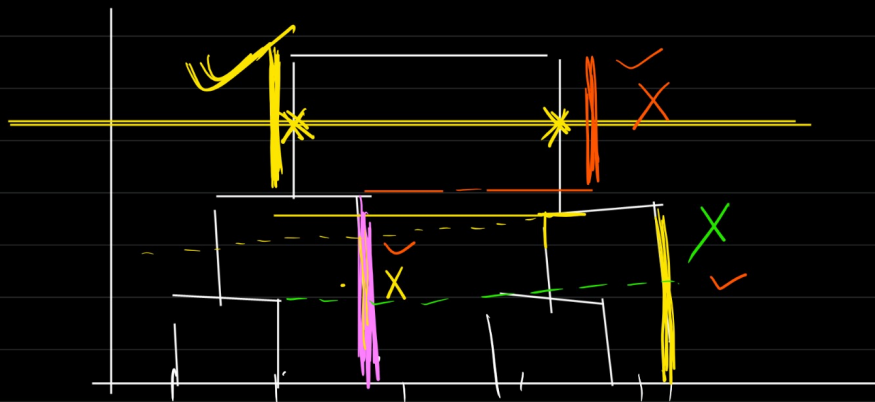
To determine  $K$ .

→ business team

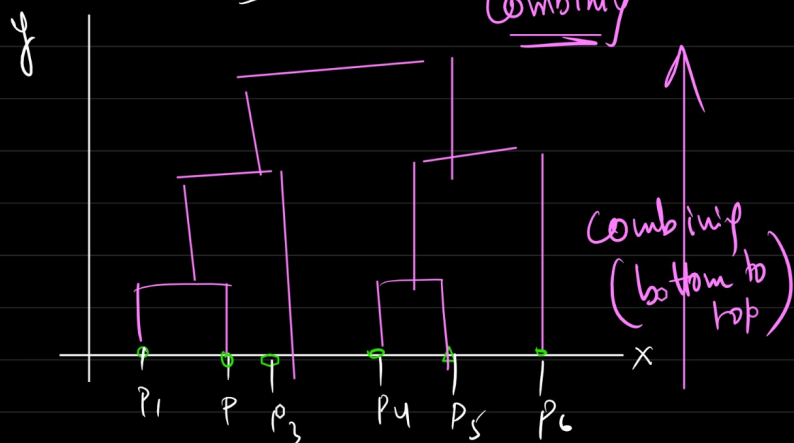
→ threshold on Euclidean distance (can be tricky)

→ We take longest vertical line of dendrogram where none of the horizontal line is passing =

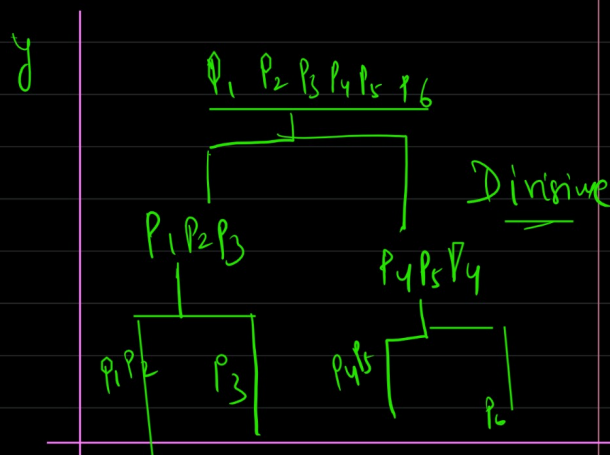
$K=2$



Agglomeration



→ Here all P's are getting combined



→ keep dividing until every individual P's

## \* kmeans Vs Hierarchical.

becomes cluster in its  
own.

- ① Kmeans works very very well for large data / Huge data.  
Hierarchical works well for small data.
- ② Kmeans  $\rightarrow$  numerical data.  
Hierarchical  $\rightarrow$  variety of data.
- ③  $K \rightarrow$  deciding  $K$  in kmeans  $\rightarrow$  using elbow.  
Hierarchical  $\rightarrow$  No centroid.