

Unsupervised learning

Name	Income	Job Profile
$X \rightarrow [x_1, x_2, x_3, \dots, x_n]$	^{n_{obj}} features/objects	

Google search

ex

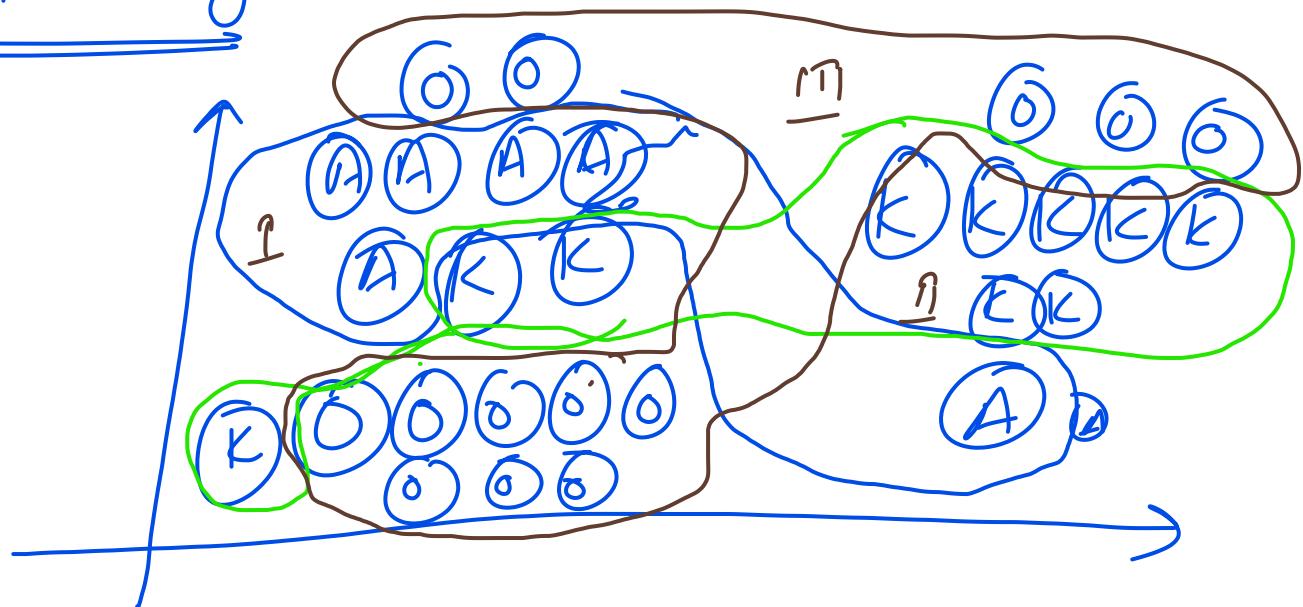
manifold

Google search ↴

ex

Security speeches.

Clustering -



Types of clustering

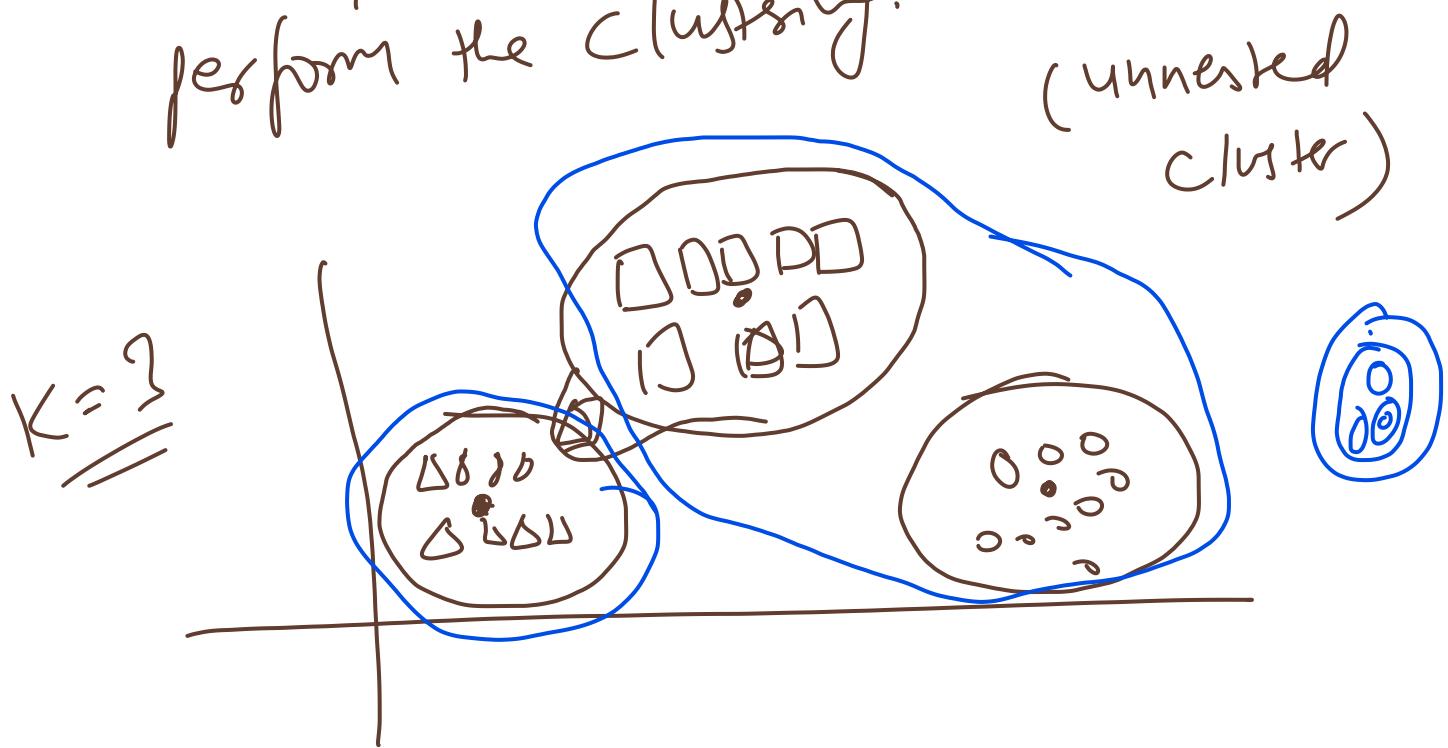
↓
Partitional
clustering

↓
Hierarchical
clustering

↓
Agglomerative

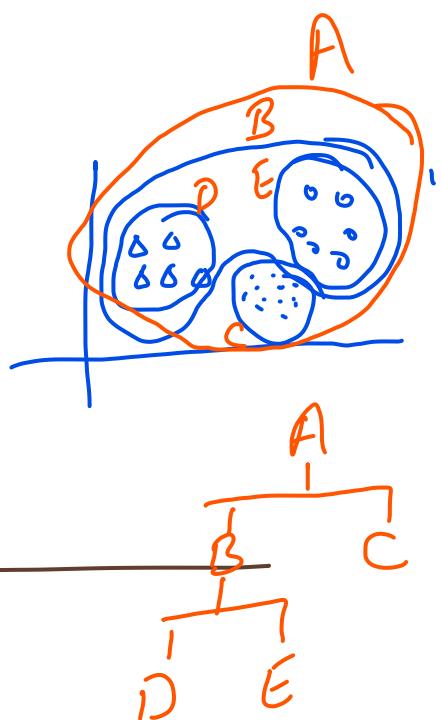
↓
Divisive

① partitioned clustering - on the basis of similarity or dissimilarity, we will perform the clustering.

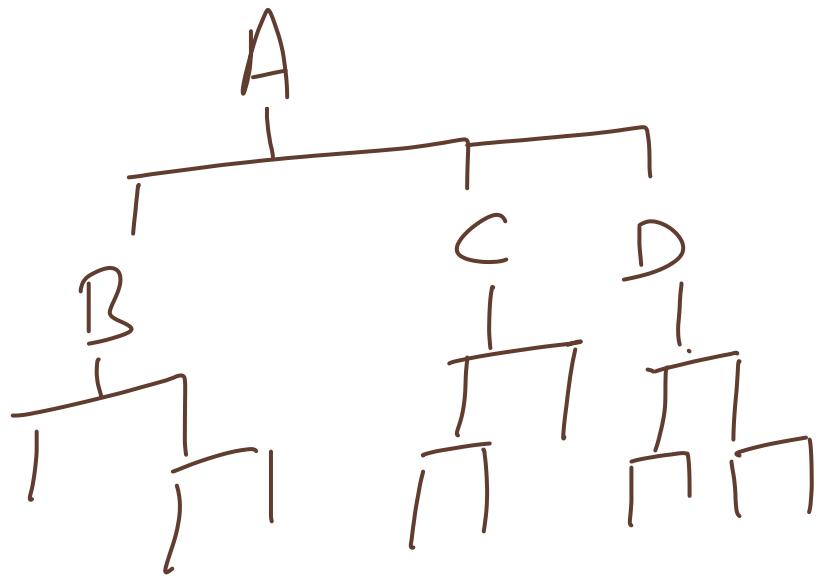


ex K-means

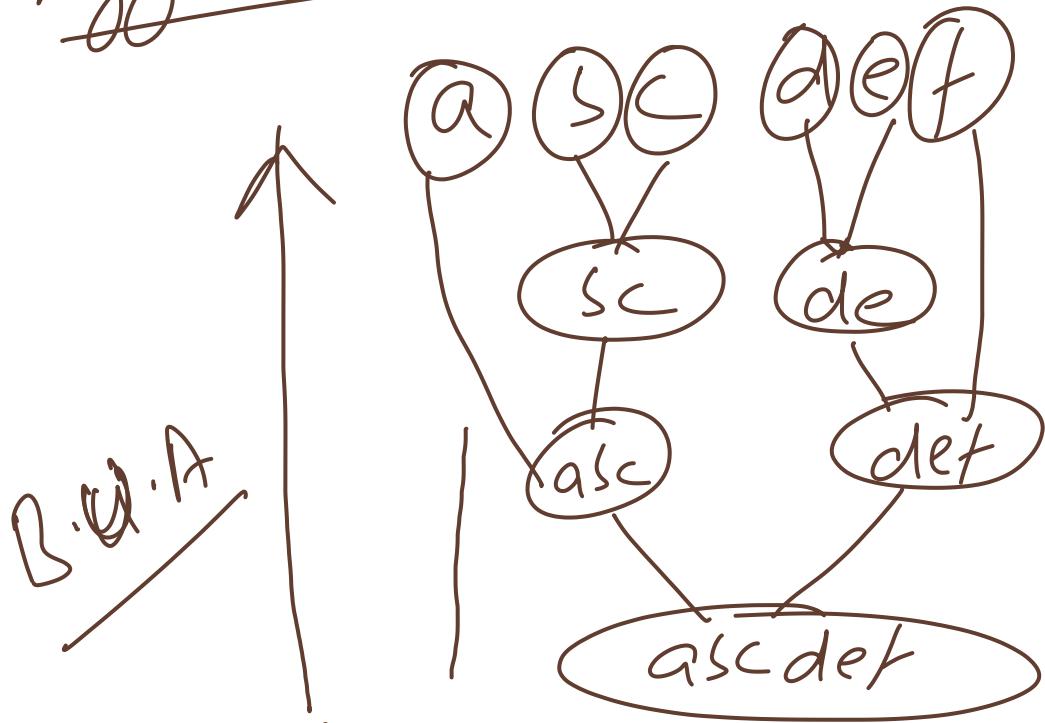
② Hierarchical clustering -
(nested clustering)



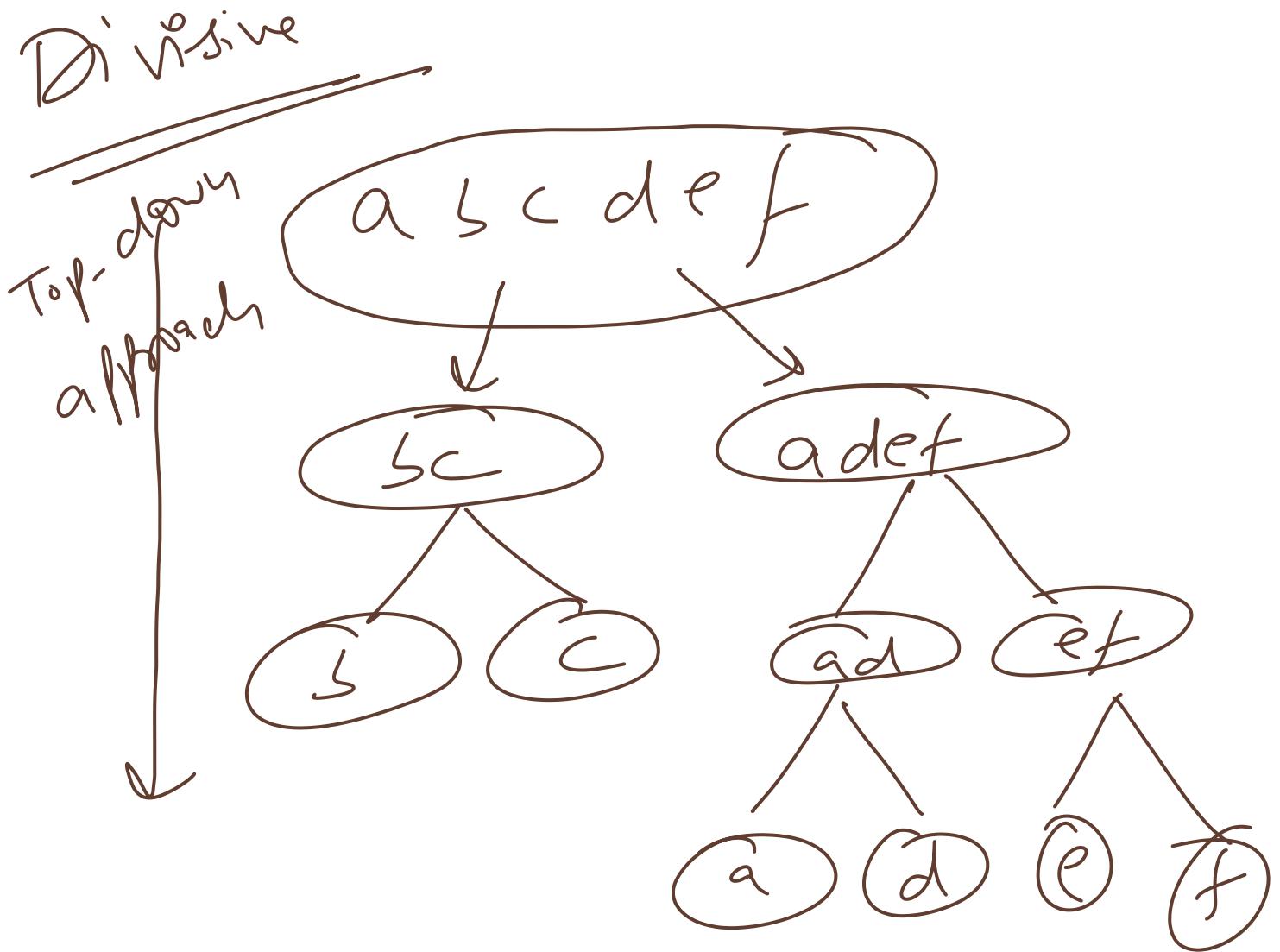
Dendrogram



Agglomerative (Bottom-up approach)

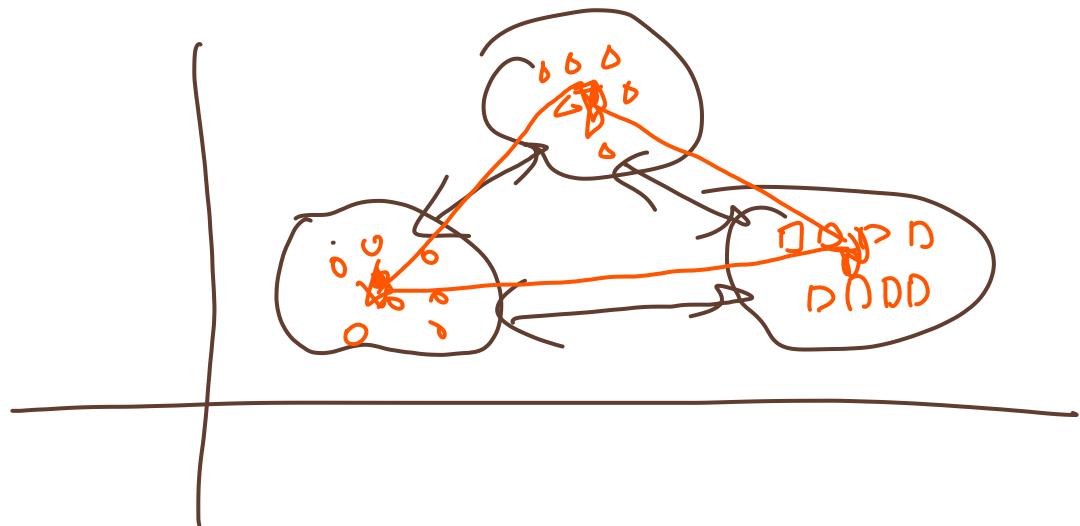


Bottom up approach

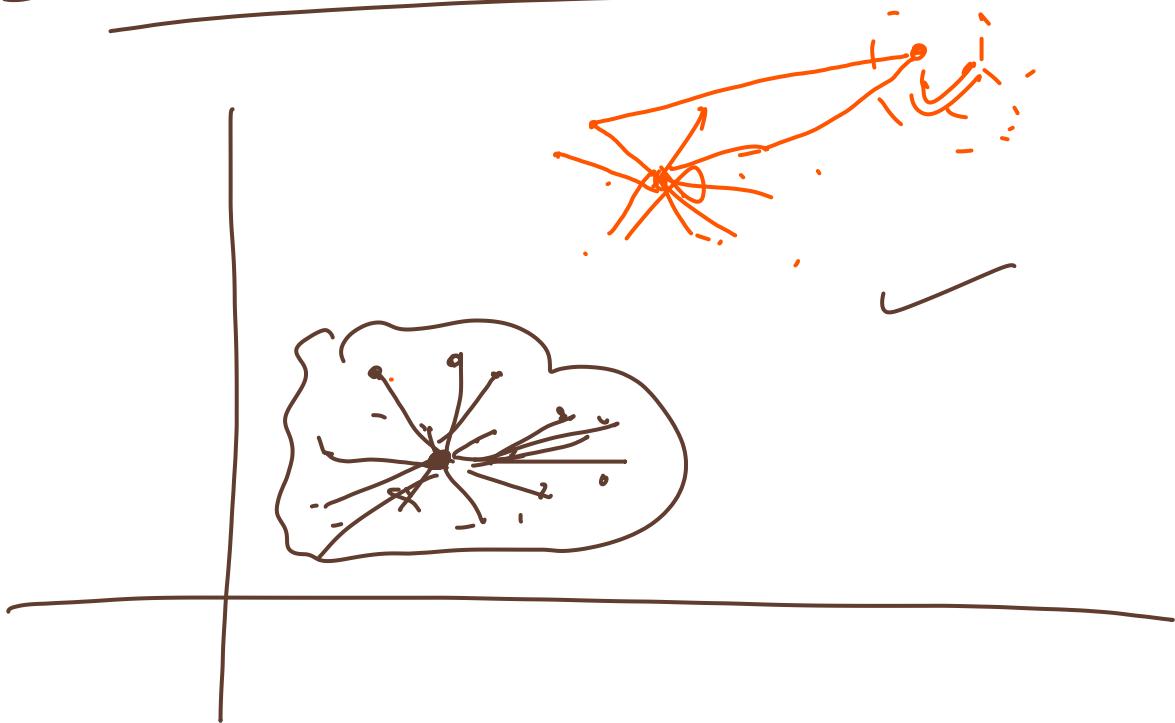


Distances in clustering -

① Inter cluster distance

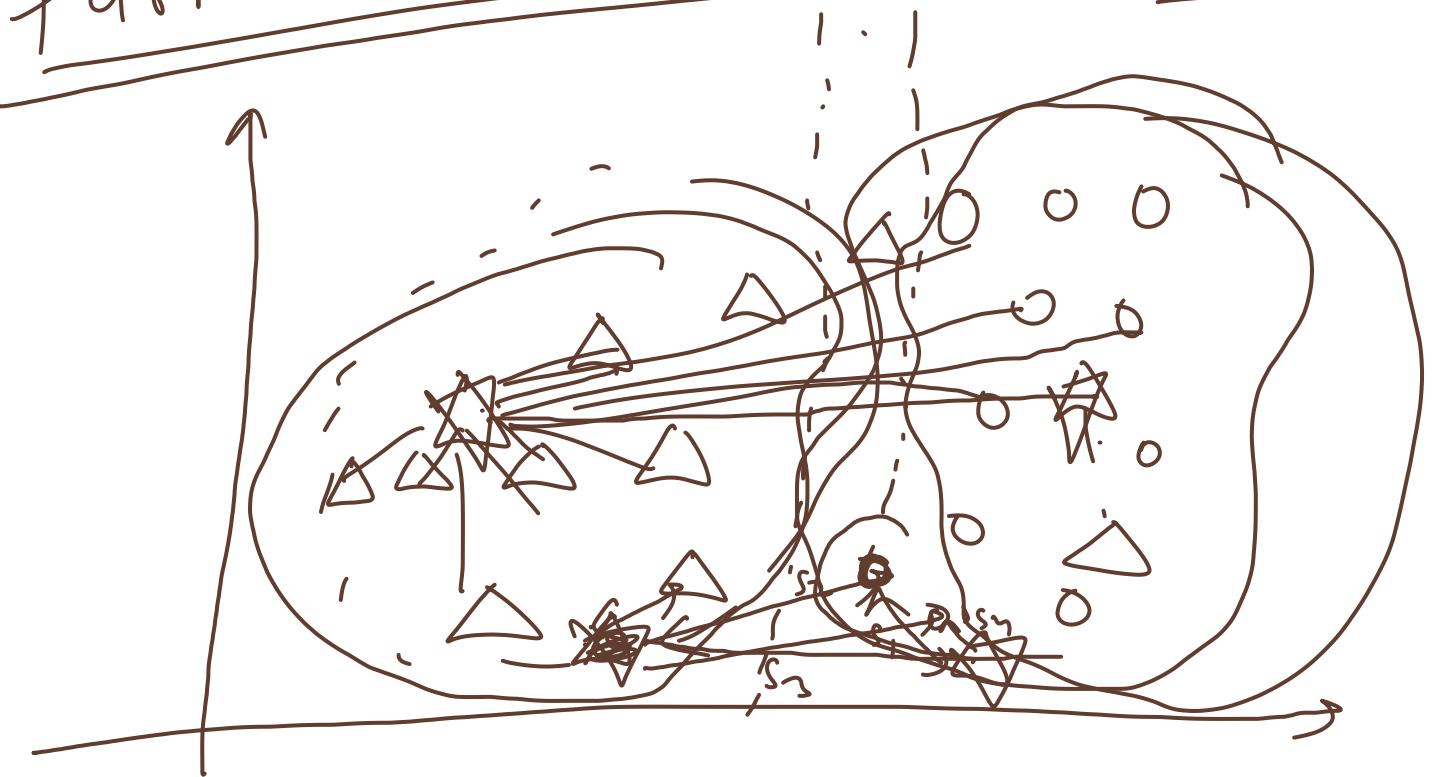


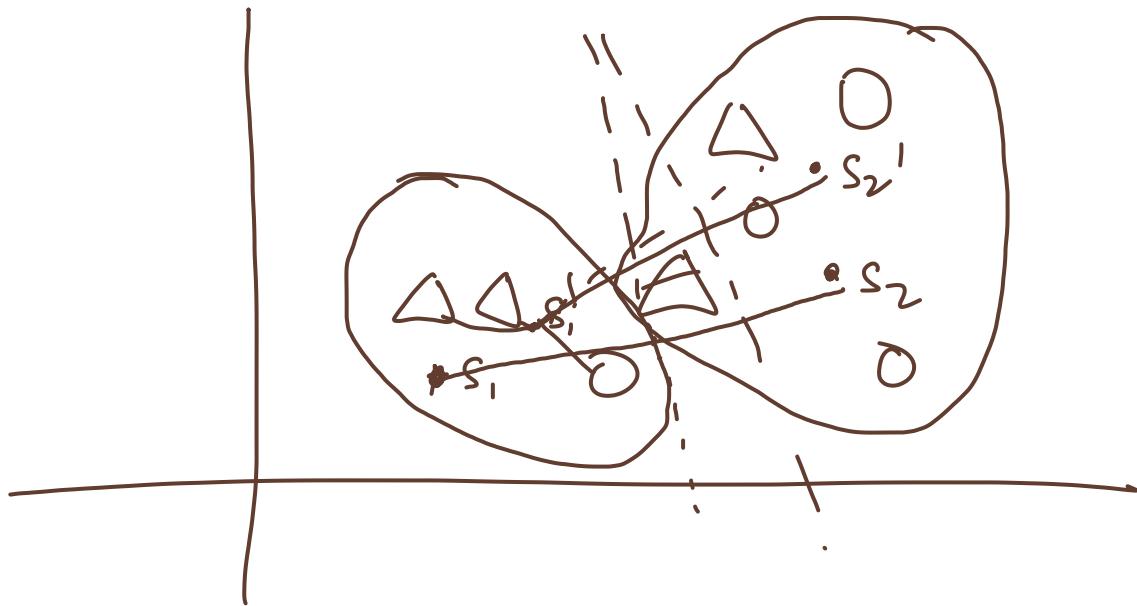
② Intra cluster distances



* Clustering is ambiguous in nature.

Partitioned clustering - (k-means)





Algorithm (K-means)

Step 1 :- Choose K points randomly.

Step 2 :- Calculate distance and Assign them to a chosen point (Cluster).

$$\therefore \text{Euclidean distance} = \sqrt{\sum_{s=1}^n (x_{is} - y_s)^2}$$

Step 3 :- Optimization criteria (stopping criteria)

SSE : sum of squared distance
(err)

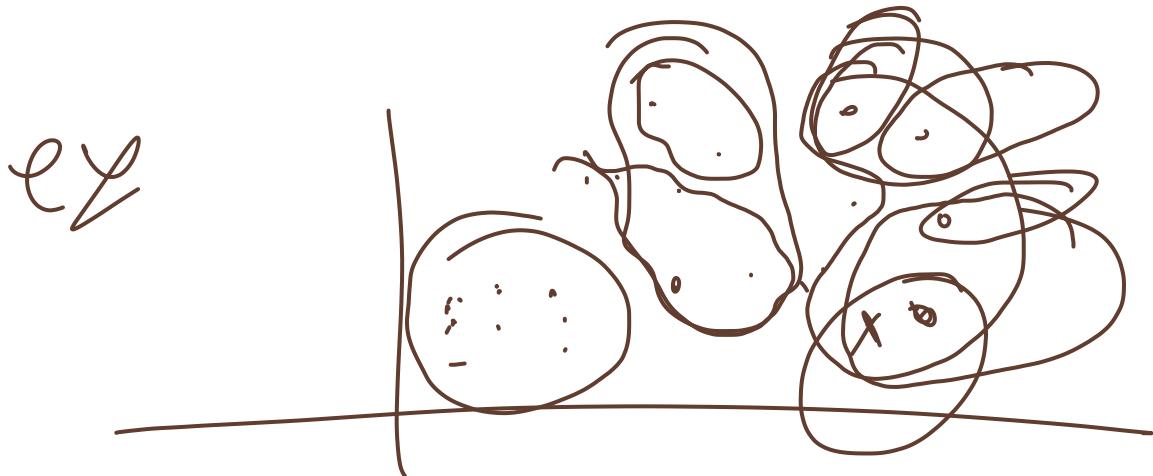
$$SSE = SSE_1 + SSE_2 + SSE_3 + \dots + SSE_K$$

$$SSE_i = \sum_{x \in S} \|x_i - c_i\|^2$$

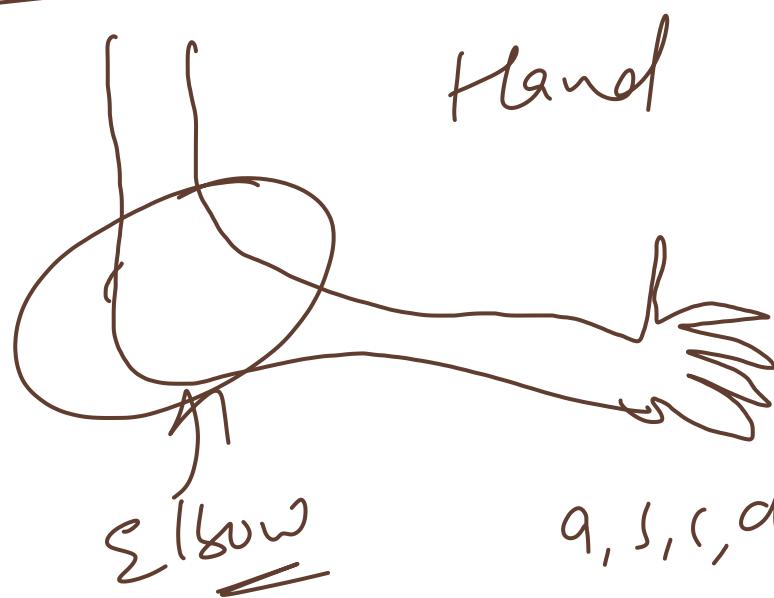
c_i : centroid

S ; cluster

Step 4 :- Recompute the cluster center and move to step ①

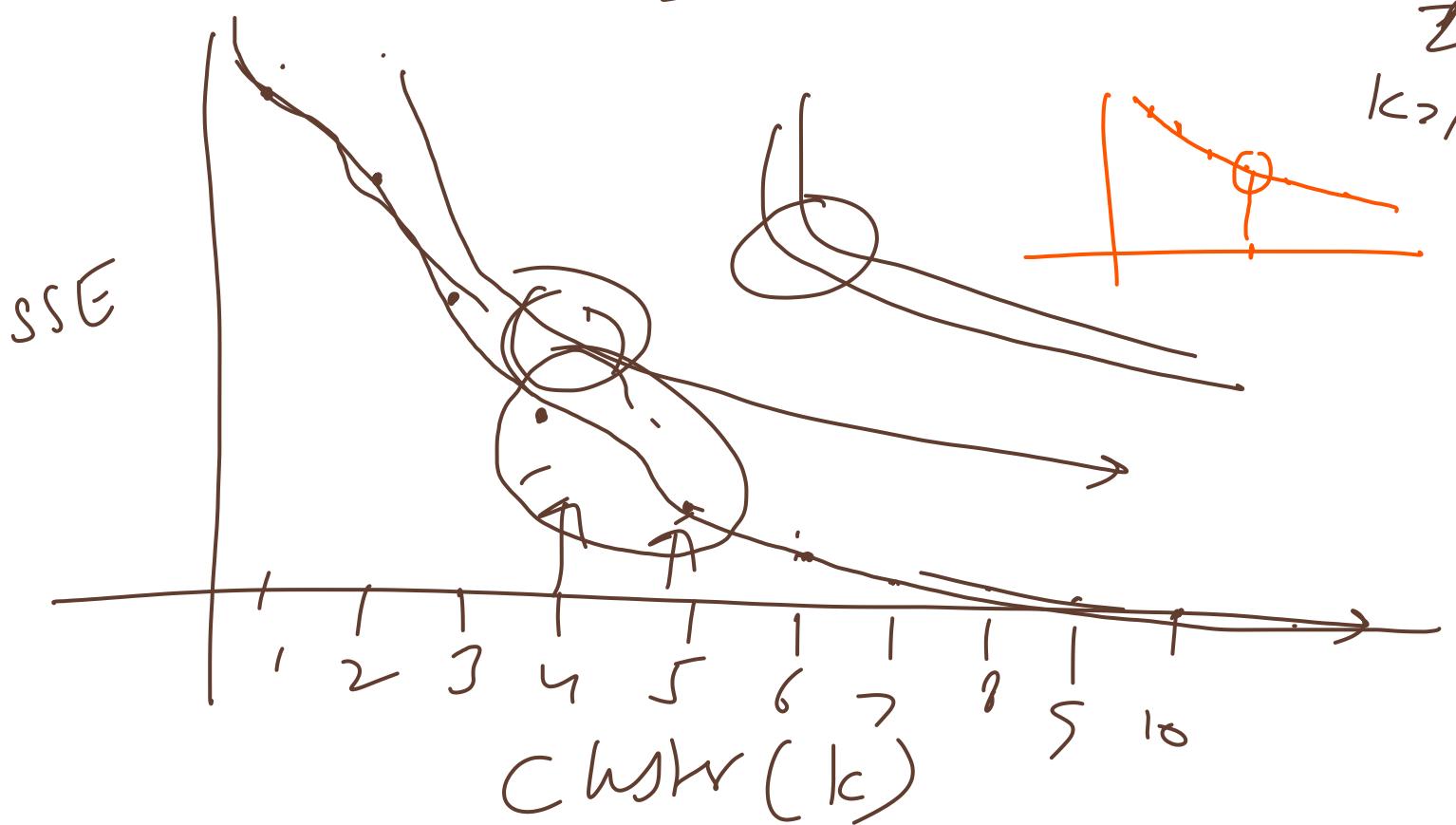


Elbow Method



$a, s, c, d, p, f, g, h, i, j$

Σ



$$X_2 \left[x, x_2, x_3, \dots \right]$$

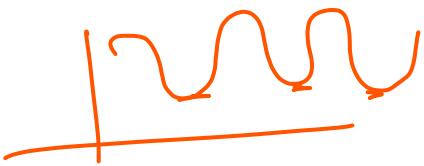
$$O/P \rightarrow S \rightarrow \{S_1, S_2, S_3, \dots, S_k\}$$

Advantage →

① fast

② Robust (easy always converge)

③ Efficient



Disadvantage -

① explicitly no of class

② may converge to local optima

(local optimum)

missed global optima

