

ML

Supervised  
(target)

Linear Regression

Logistic Regression

Unsupervised  
(no target)

Clustering

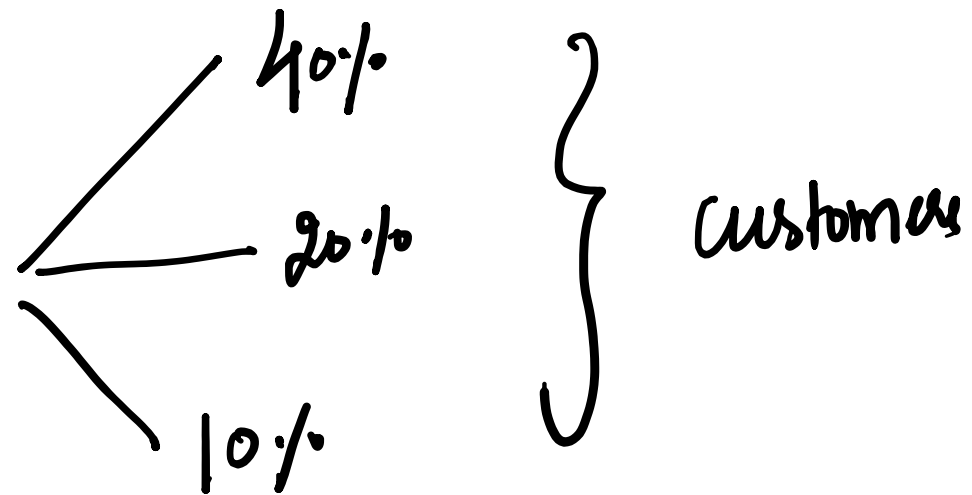
- K-Means
- Hierarchical

→ Need of Clustering:-

Amazon → Growth team

↓  
Operations team

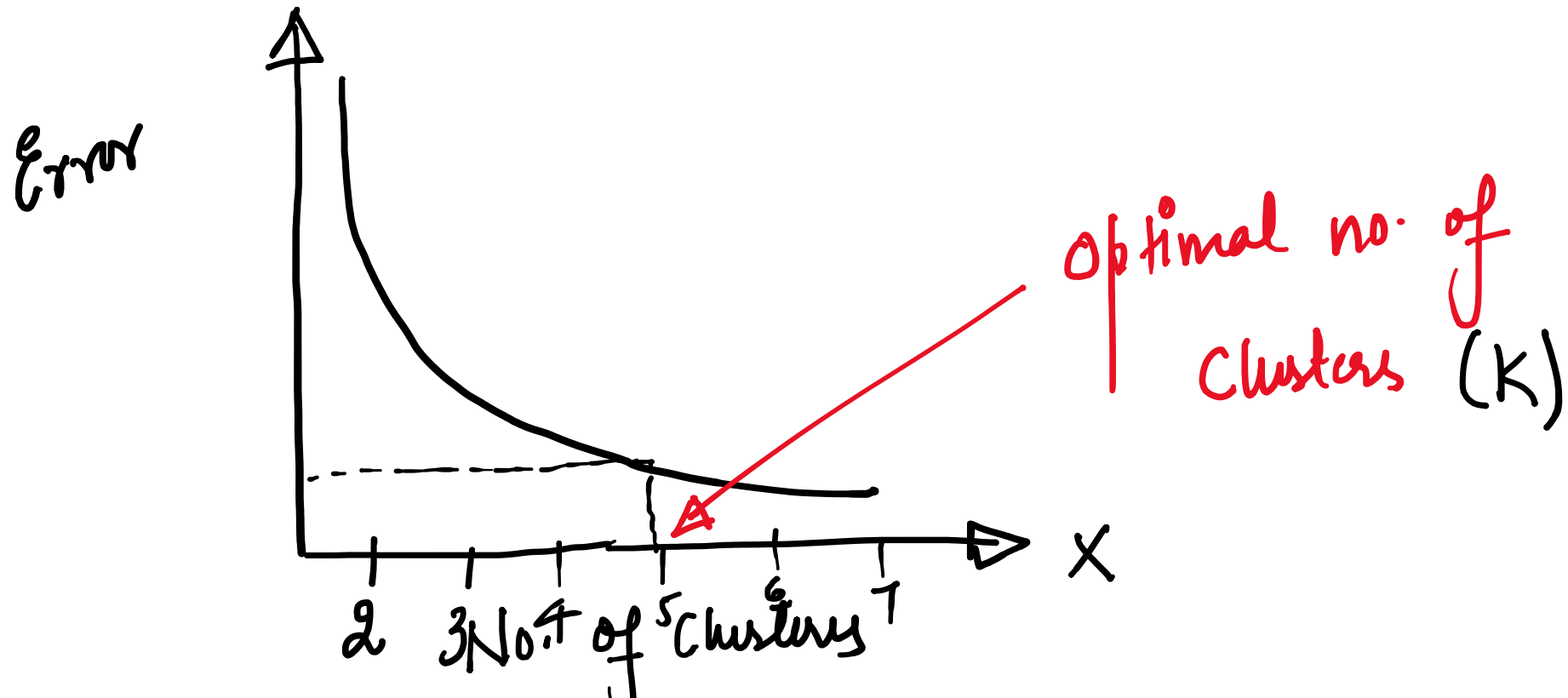
- ① → Distribute coupons randomly
- ② → Segment the customers into 3 segments basis their historical activity on platform.



→ ~~K~~ Means Clustering :- \* No. of clusters to be known  
beforehand (90-95%)

↓  
No. of clustering

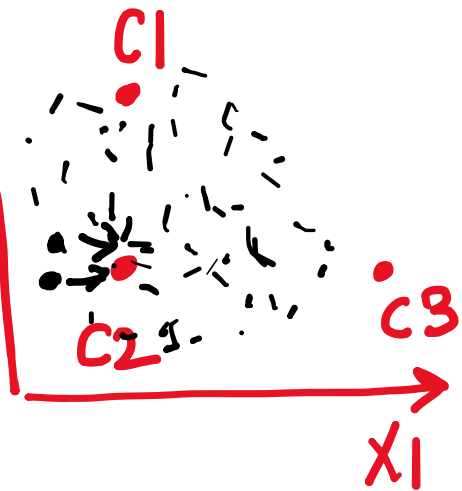
→ Elbow method :-



→ Algorithm :-  $K=3$

✓ ① Random Initialisation : Randomly initiate  $K$  centroids in vector space.

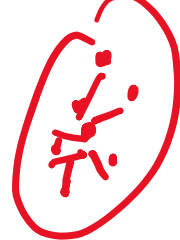
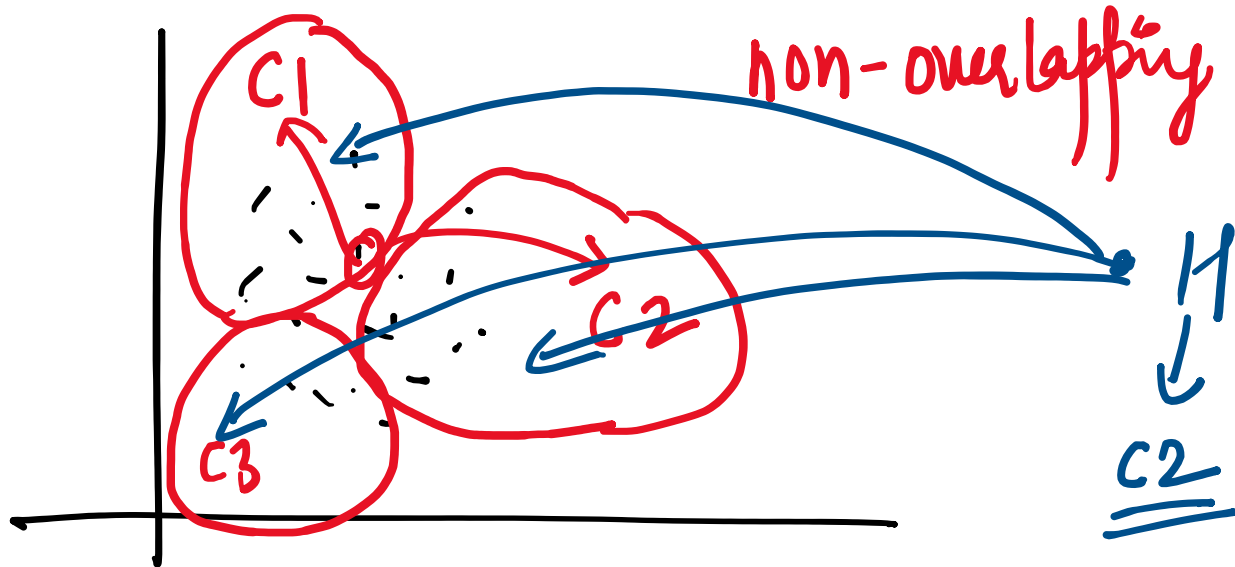
✗ ② Assignment :- Each data point is assigned to its nearest centroid basis Euclidean distance.



✗ ③ Update :- New value of centroid will be mean of all data points assigned to it.

✗ ④ Repetition :- Repeat step ②, ③ until convergence.

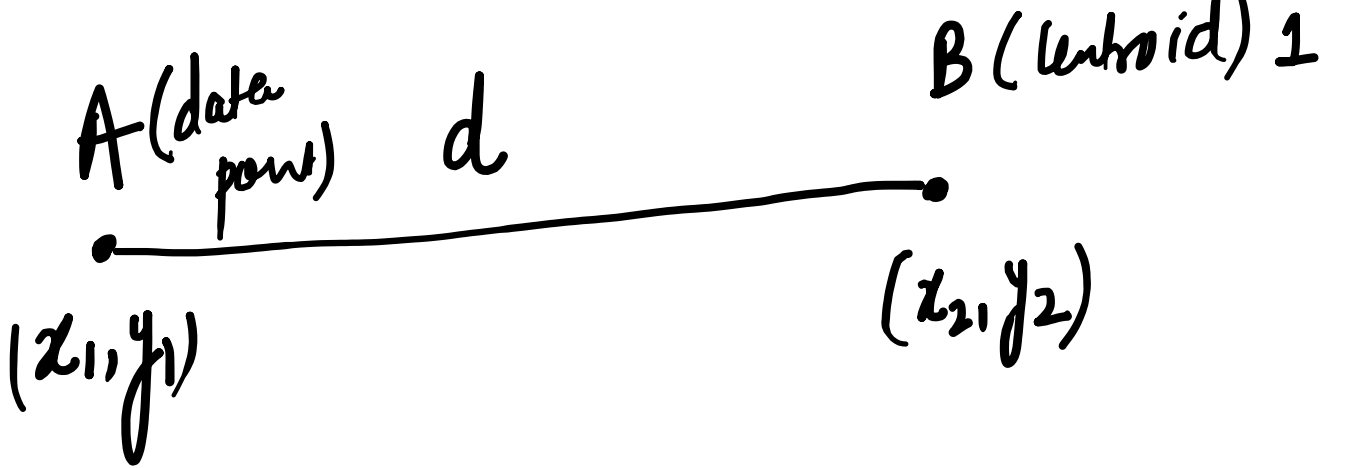
$K=3$



maximise  
intercluster distance  
minimise  
intracluster distance

→ K-Means++ :- Modified version of K-means

→ Euclidean Distance :-



A (data point)  $(x_1, y_1)$   $d$  B (centroid) 1  $(x_2, y_2)$

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

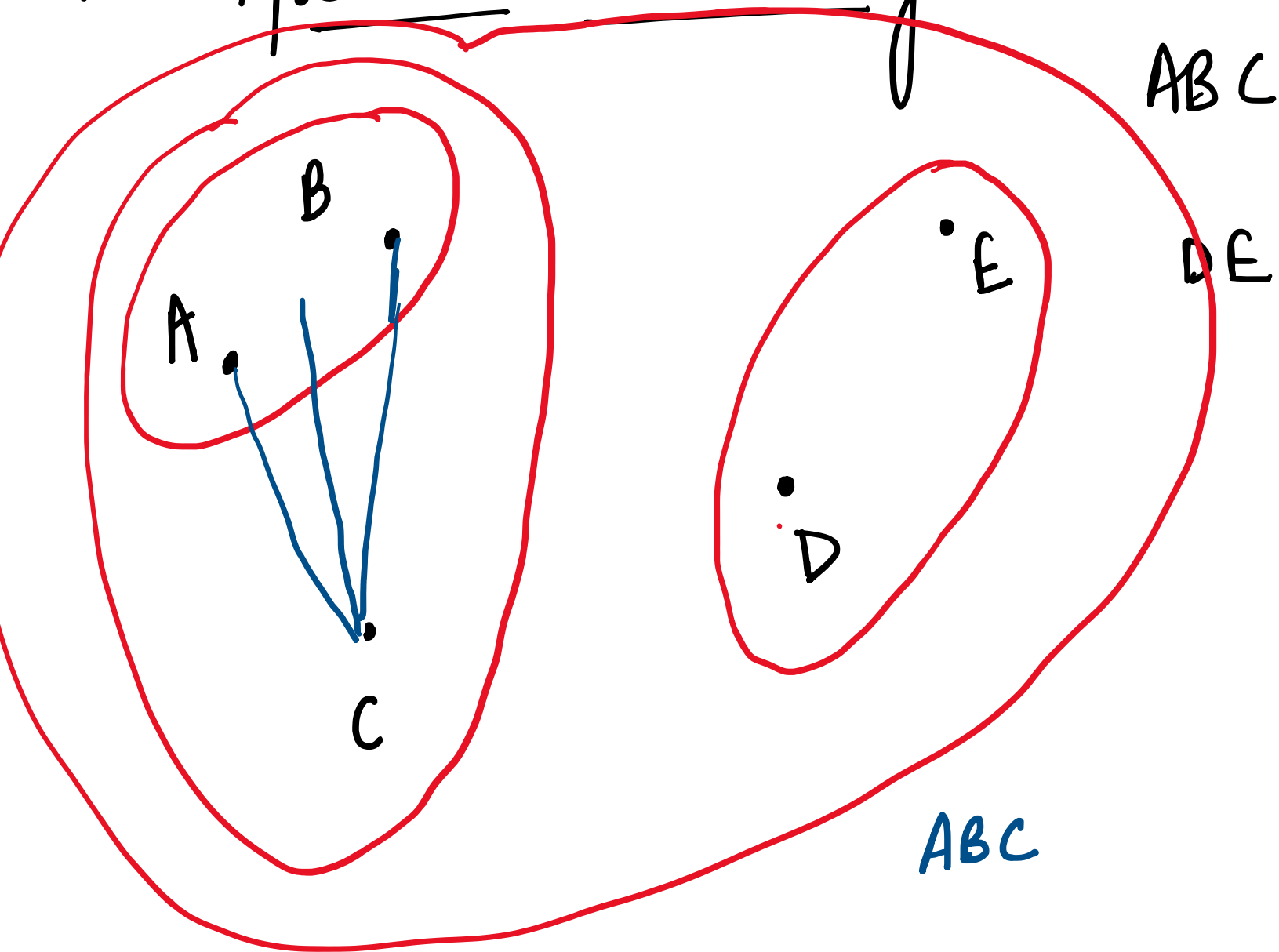
$$c_2 = \frac{A+B}{2}$$

A  
c2 B

$$c_1 = \frac{D+C}{2}$$



→ Hierarchical clustering :-

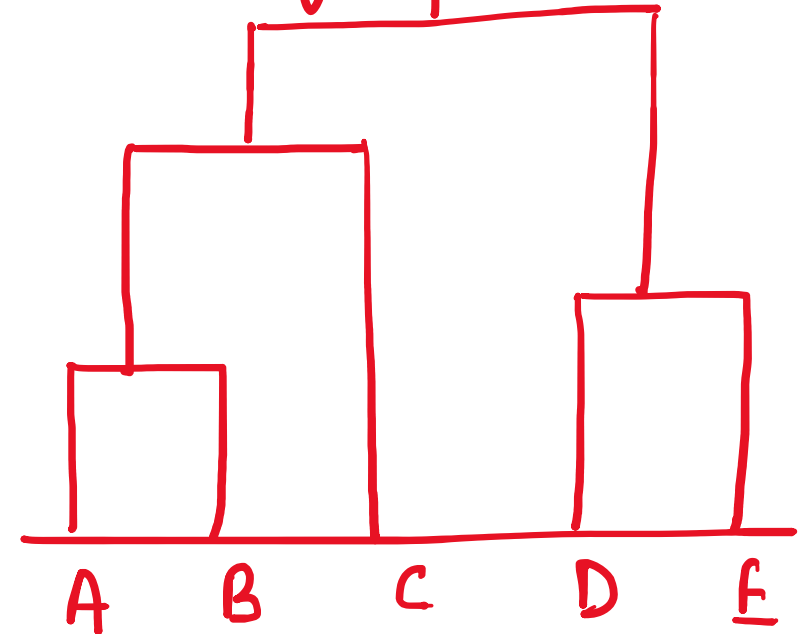


ABC

DE

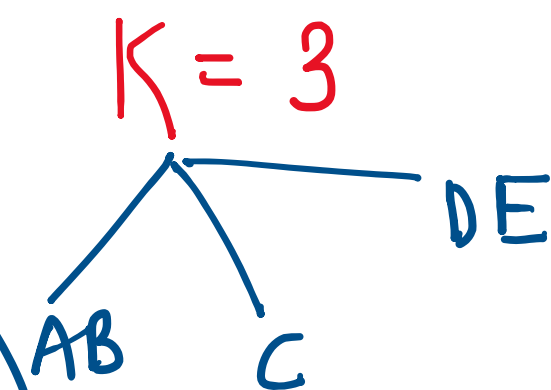
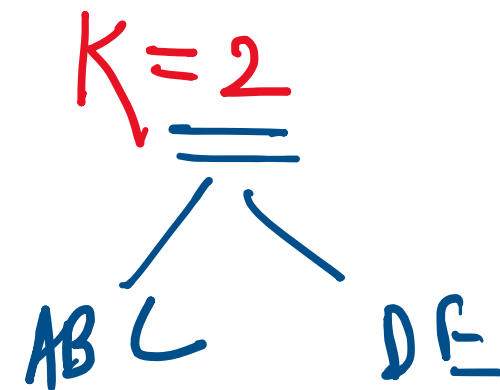
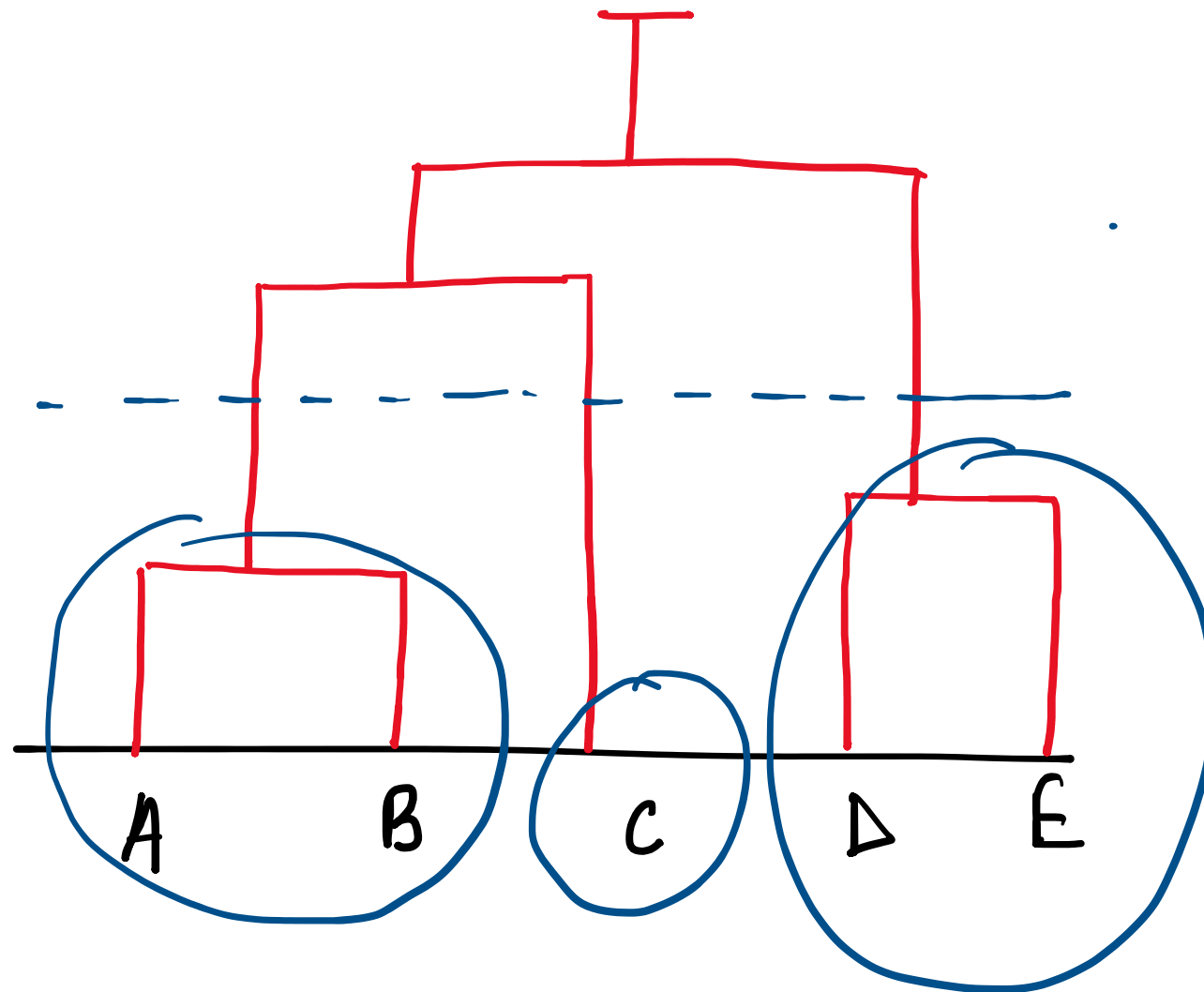
ABCDE

Dendrogram



→ How to make clusters from dendrogram?

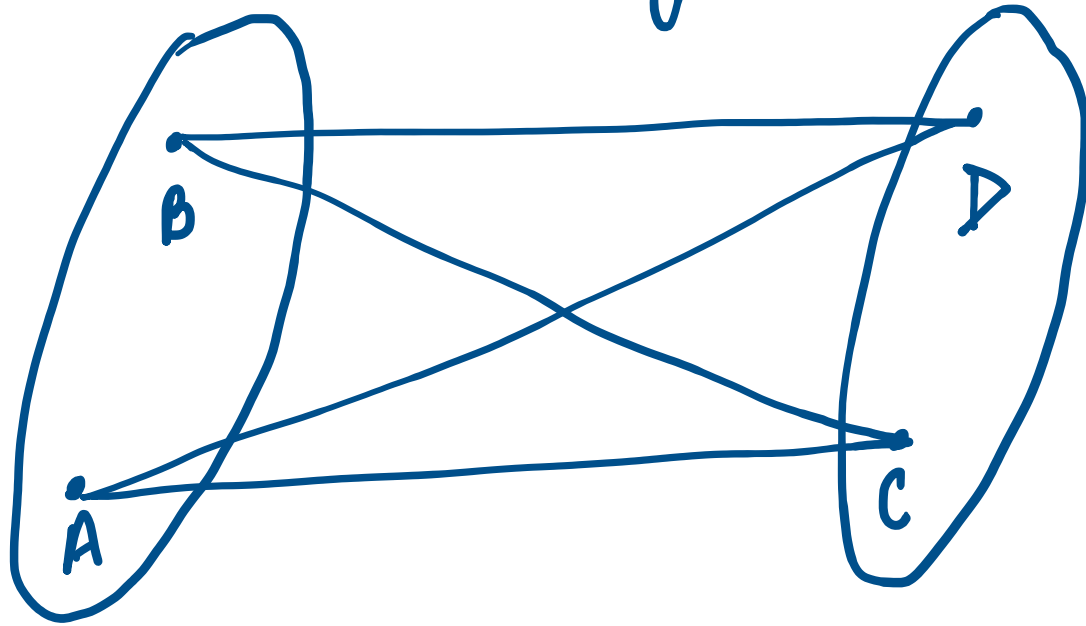
i





→ Linkages:

- Complete
- Single
- Average



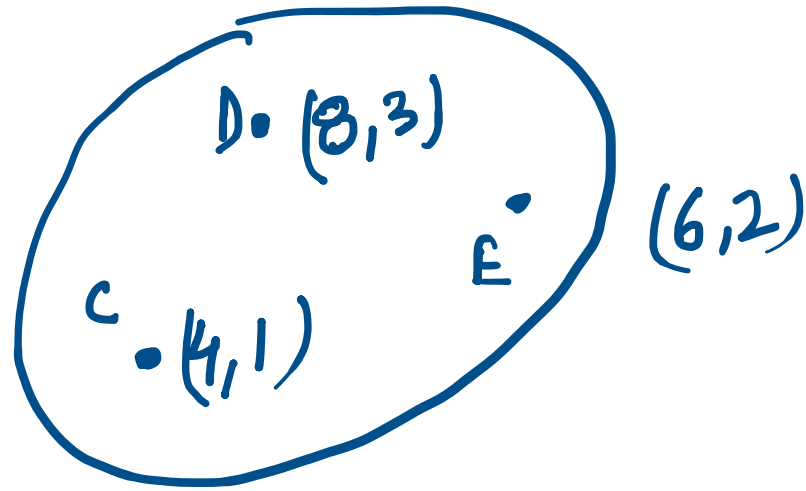
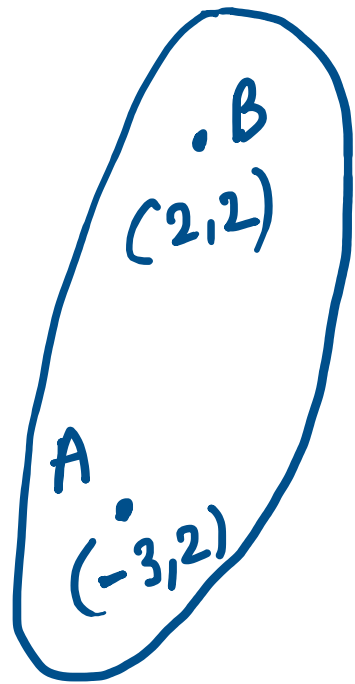
ABCD

Complete → (AD, AC, BD, BC) Max

Single = (AD, AC, BD, BC) Min

Average = (AD, AC, BD, BC) Average

Ques



Complete linkage  $\Rightarrow$

ABCDE  $\rightarrow$   
 $\downarrow$  Max  
AD

$$AD = \sqrt{11^2 + 1^2}$$

$$BD = \sqrt{6^2 + 1^2}$$

$$BC = \sqrt{2^2 + 1^2} = \sqrt{5} = 2.23$$

$$AC = \sqrt{7^2 + 1^2}$$

$$AE = \sqrt{9^2 + 0^2}$$

$$BE = \sqrt{4^2 + 0^2} = 4$$