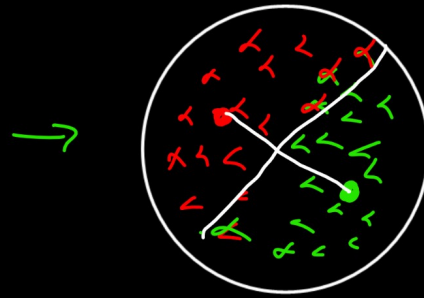
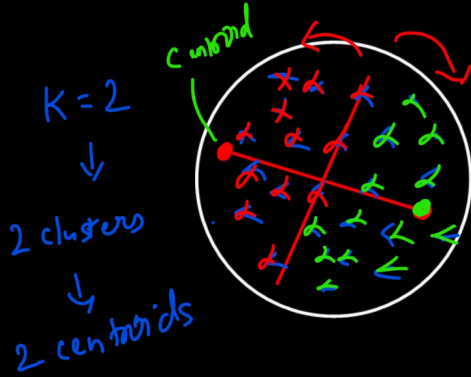


UNSUPERVISED LEARNING → Train the machine with unlabelled data (unknown data)

↓
(1) K-Means clustering (2) Hierarchical clustering

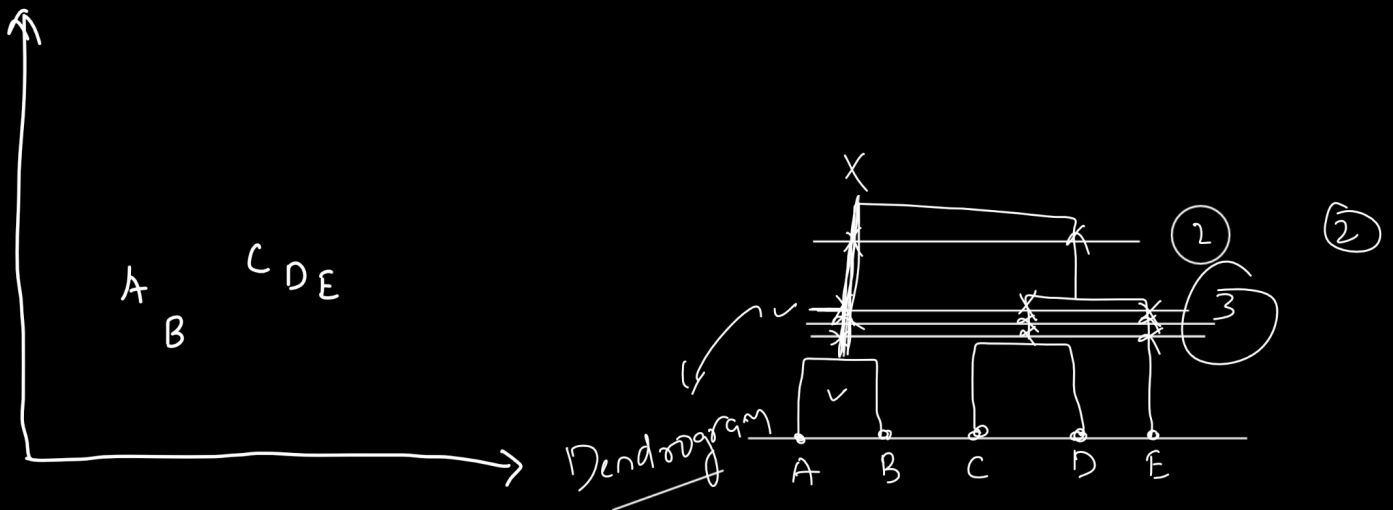
K-Means Clustering → K = No. of clusters



Continue
till
there is no change

Hierarchical clustering — (1) Divisive clustering (Top to Bottom)
(2) Agglomerative clustering (Bottom to Top)

(1) Agglomerative clustering: → Dendrogram → K-value



5 samples → 5 iterations

1st Iteration → Assume all data points as individual clusters → 5 clusters

2nd iteration → 4 clusters (A, B, C, D, E)

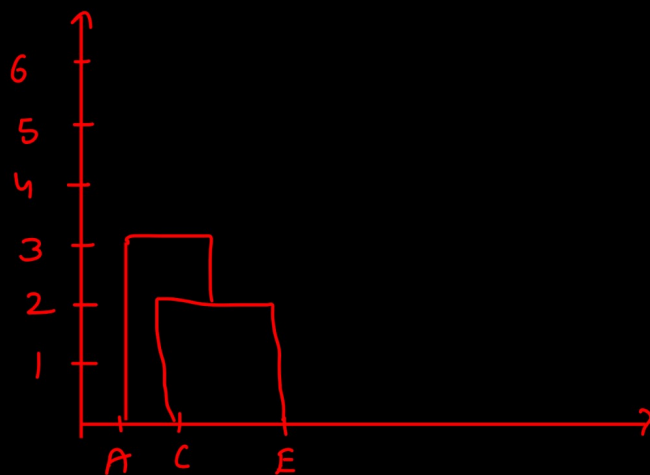
3rd iteration → 3 clusters (A, B, C, D, E)

4th iteration → 2 clusters (A, B, C, D, E)

5th iteration — 1 cluster (A, B, C, D, E)

Distance Matrix

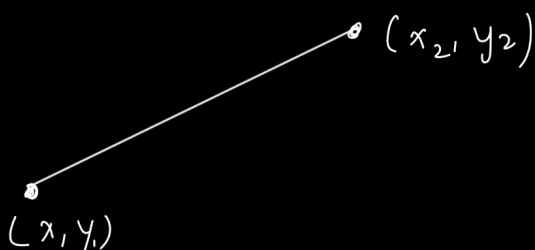
	A	B	C	D	E
A	0	9	3		
B	9	0			
C	3	7	0		
D	6	5	9	0	
E	11	10	2	8	0



Distance Formula — ① Euclidean — (least)
② Manhattan

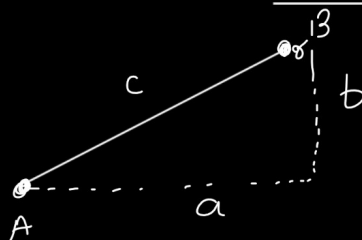
$$\text{Euclidean} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

$$\text{Manhattan} = |x_2 - x_1| + |y_2 - y_1|$$



Euclidean = c

Manhattan = a + b



Linkage — Avg, Single, Multiple
↓ ↓ ↓
Avg Min Max

	A	B	CE	D
A	0	9	<	<
B	9	0	<	<
CE	3	7	0	<
D	6	5	8	0

A-CE → $\begin{bmatrix} A-C & 3 \\ A-E & 11 \end{bmatrix} \rightarrow 3$
B-CE → $\begin{bmatrix} B-C & 7 \\ B-E & 10 \end{bmatrix}$
CE-D → $\begin{bmatrix} C-D & 2 \\ E-D & 8 \end{bmatrix}$

	A-CE	B	D
A-CE	0	7	
B	7	0	
D	2	8	0

B-A-CE $\begin{bmatrix} B-A & 9 \\ B-C & 7 \end{bmatrix}$
B-E-10