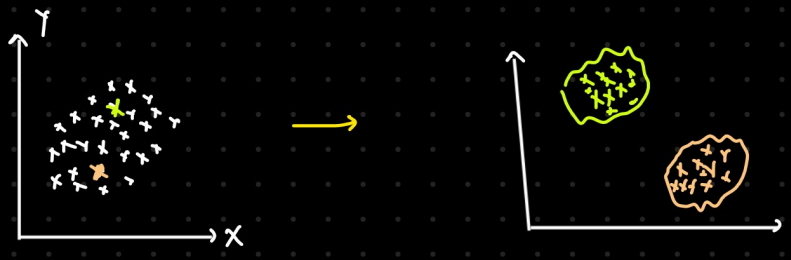
1. K-Means → K-Means++
2. hirarchical.
3. DBScan

1 K-Means. | K-Means++ Updated version of the K-means.

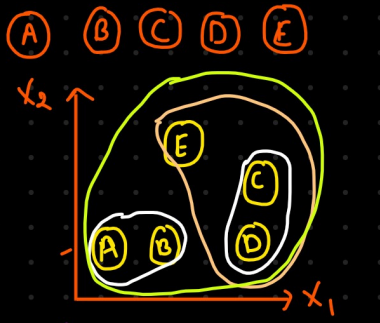
1. random initialization of Centroid.
2. find out the dist- to all the point- and make cluster (min dist)
3. Update the Centroid

K-Means algo is Centroid based algo

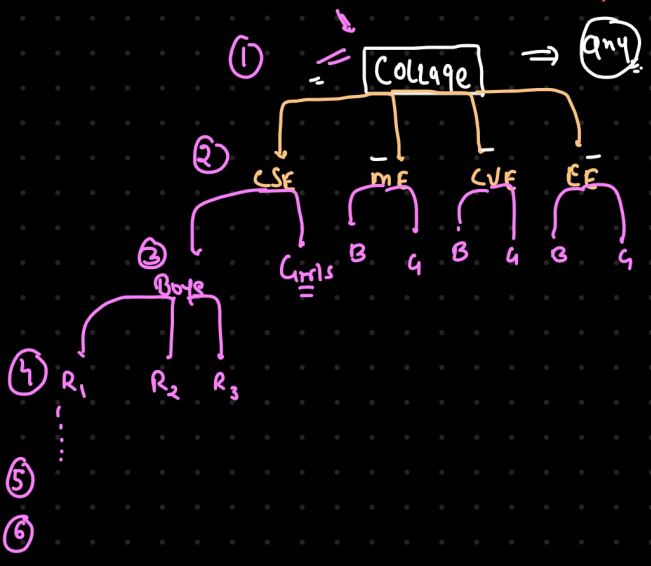


K = 2.
K = 3
K = 4

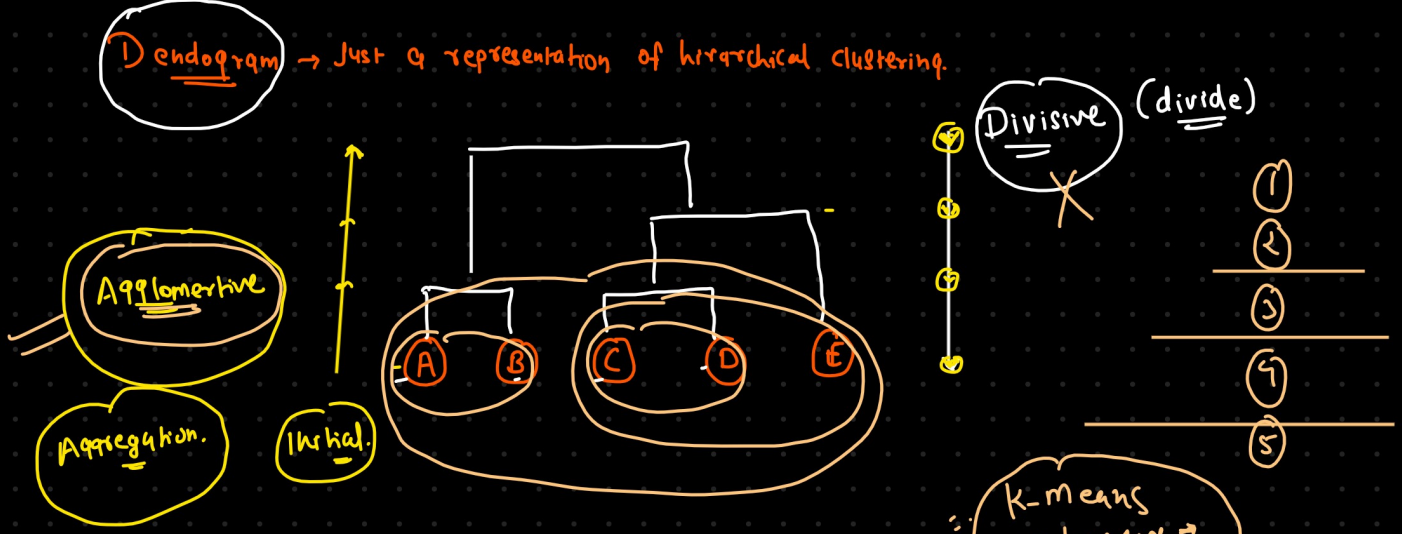
2 hirarchical clustering.



Point matrix



① Dendogram → Just a representation of hirarchical clustering.

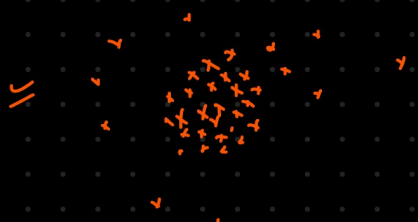


K-means
Agglomerative →
DBScan

Agglomerative →
① each Point is cluster
② Bottom to top approach.
③ Combining all the Point as a single cluster.

③ DBSCAN — Density based spatial clustering with application with noise

- Density based approach.



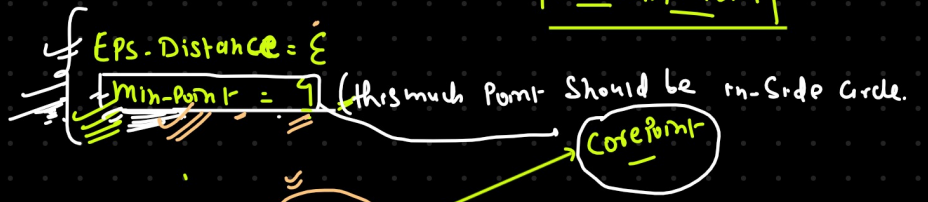
K-means
hirarchical
DBScan.

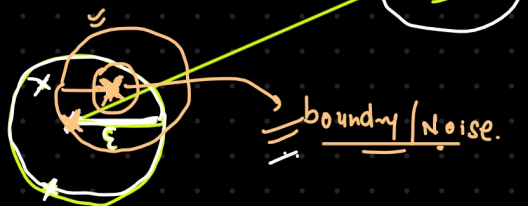
hyperparameter →
① eps dist-
② min Point-

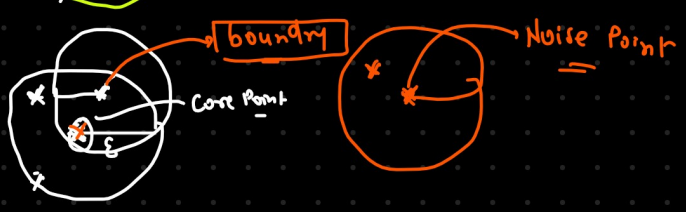
Cluster ←
(outlier)

ε
① ~~eps distance.~~
② ~~Core Point~~
③ Border Point-
④ Noise Point-

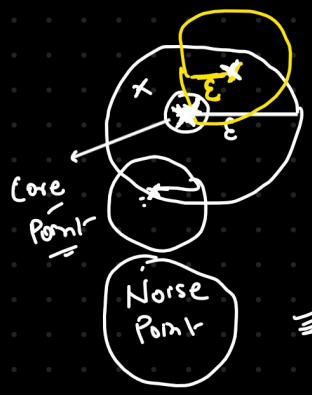
~~Minimum - Point~~







- (1) eps distance.
- (2) mini-Point to be Consider in a Circle.
- (3) Core Point-
- (4) border Point-
- (5) Noise Point -



(1) min Point = 4

Y | N

It is not Core Point

(2) Border Point Condition → it should be Neighbor of the Core Point

(1) K-means.
(2) Hirarchical. → Point matrix.
(3) DBScan.

How many Point → 5 Point (P1) (P2) (P3) (P4) (P5)

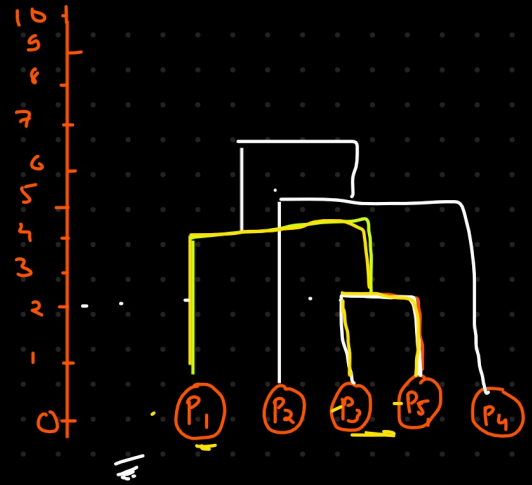
find dist. b/w the Point and creat cluster

$P_1(4,5)$ → E.D →
$P_2(6,7)$
:

	P_1	P_2	P_3	P_4	P_5
P_1	0				
P_2	9	0			
P_3	3	7	0		
P_4	6	5	9	0	
P_5	11	10	2		0

min-dist b/w Point-

P_i	P_2	$[P_3, P_5]$	P_4
P_1	0		
P_2	9	0	
$[P_3, P_5]$	3	7	0
P_4	6	5	8	0



$d(P_1, [P_3, P_5])$

$\min [d(P_1, P_3), d(P_1, P_5)]$

$\min [3, 11]$

$= 3$

$d(P_2, [P_3, P_5])$

$[d(P_2, P_3), d(P_2, P_5)]$

$\min [7, 10]$

$\Rightarrow 7$

	$[P_1, P_3, P_5]$	P_2	P_4
$[P_1, P_3, P_5]$	0		
P_2	7	0	
P_4	6	5	0

$D[P_2, [P_1, P_3, P_5]]$

$\min \; D[(P_2, P_1), (P_2, P_3), (P_2, P_5)]$

$= 7$

① K-means.

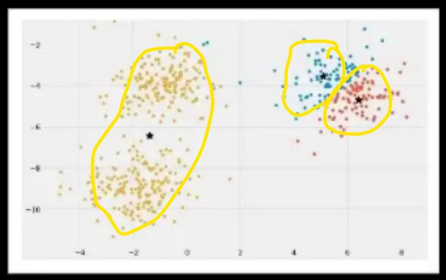
&

K-means ++

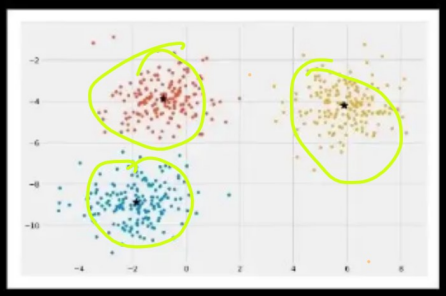
② hirarchical. (Mathmatical) Point matrix

(Agglomerative)

③ DBScan.



good ①



good ②

① If all about Centroid initialization.

K-means → Randomly ✗ In many cases.

↓

K-mean ++

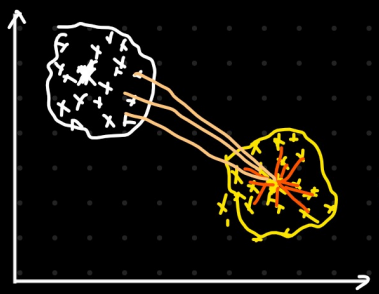
Probabilistic

C $\xrightarrow{d}$ x_i

farest

Clustering

1. Intra cluster Dist (inside same cluster)
2. inter Cluster Dist (Between a cluster)



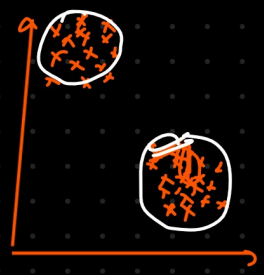
Dunn index

$$D = \frac{\max d(i,j)}{\max d(k)}$$

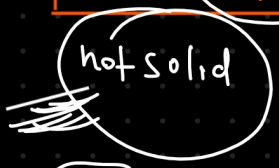
$$= \frac{\text{maximum inter cluster dist-}}{\text{max intra cluster dist-}}$$

$$= \frac{\uparrow \text{as much as possible}}{\downarrow \text{as low as possible}}$$

more solid.



more solid



$$\text{my best dunn index} = \frac{\uparrow \max(\text{inter})}{\downarrow \max(\text{intra})}$$

1. do you know kmeans
2. diff kmean v/s kmeans++
3. can you explain hirarchical clustering (math intuition)
4. DBScan.
5. Why we should use DBScan.
6. diff b/w k-means v/s hirarchical
7. How to evalute k-mean or kmeans++
8. Shilhoutee Score.
9. Dunn index.
10. inter cluster dist- and intra cluster.
11. WSSS.