Sir.

30/4/22 K-means :-

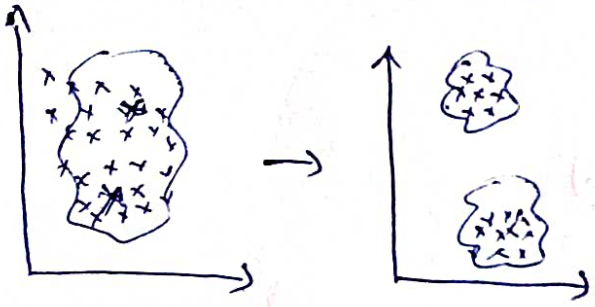
Hierarchial clustering [Point Matrix]

1- Random initialization of centroid.

2. find out distance to all the point and make cluster [min dist]

3. Update the centroid.

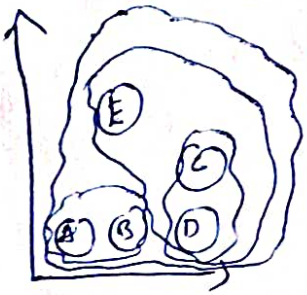
K-means algorithm is centroid based algorithm.

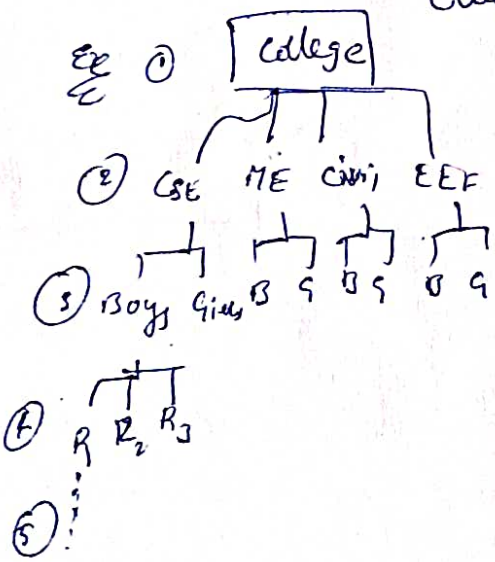


Hierarchial clustering :-

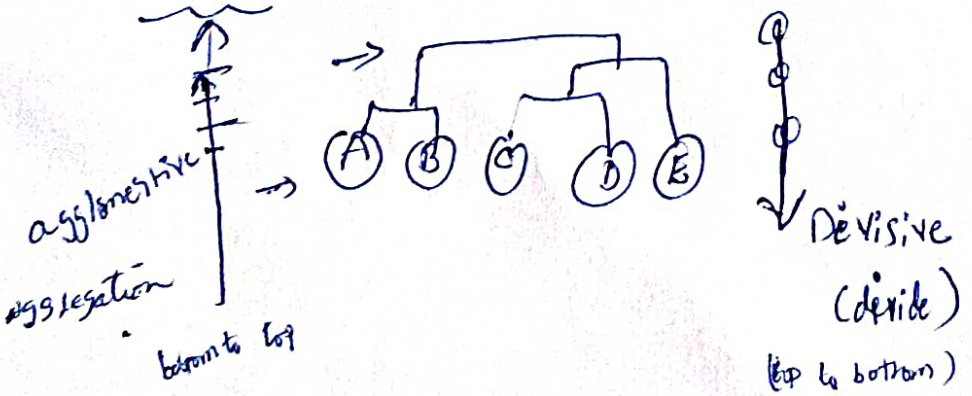
each & every posibility w.r.t clustering =

(A) (B) (C) (D) (E) → each & every possibility to make a cluster.





Dendogram :- Just a Representation of hierarchial clustering



Agglomerative :- 1. Each Point is cluster
2. Bottom to top approach
3. Combining all the Point as a single cluster.

(1) (2) (3) (4) (5) (6)

K means
Agglomerative
DB scan

DB scan :- Density Based Spatial Clustering of applications with noise

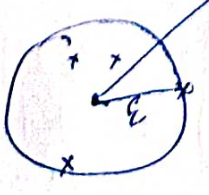
Density Based Approach

1. Epsilon distance
2. Core Point
3. Boder point
4. Noise Point
5. Minimum Point to be consider in a circle.

} based on these Points, we need to create the cluster

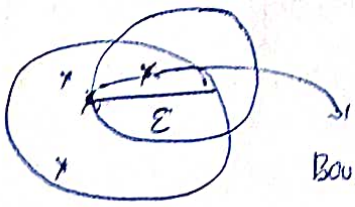
(Outlier)

1. Eps. distance = ε
min-Point = 4 (this much point should be in-side circle (core point) min point

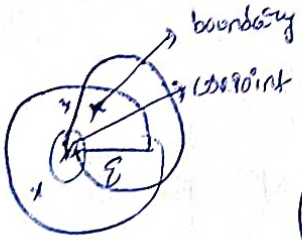


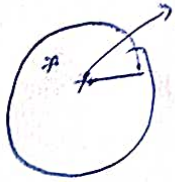
A Point is a Core point if there are at least min Pts number of Points in its surrounding area

Hyper Parameter → 1. EPS dist
2. min Point

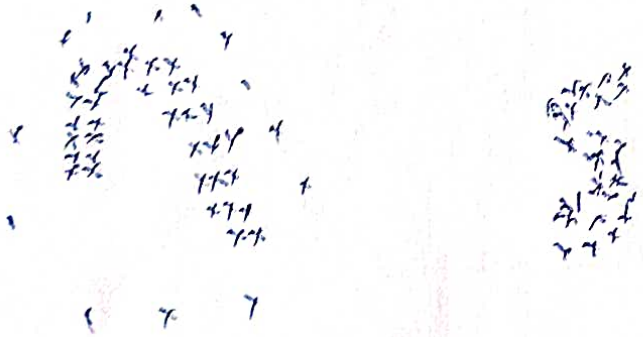


if any point, which is neibour to core point is called Border Point.





Eg



each & every point will treat as core point
epsilon distance, hyper parameter.

Hierarchical clustering :-

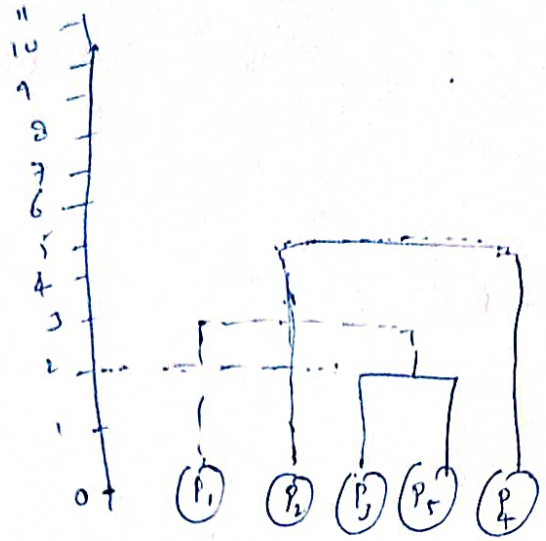
5 Point →

find out the distance btw the point and create clusters.

	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 Point

min dist Point = 2.



	P_1	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

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

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

$\min [3, 11]$

$= 3$.

Here min distan is with, $[P_1, [P_3, P_5]]$

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

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

$\min [7, 10]$

$= 7$

	(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

$D\{P_2, (P_1, P_3, P_5)\}$

$[(P_2 P_1), (P_2 P_3), (P_2 P_5)]$

* Diff btn K-means, & K-means++

↓

It's all about Centroid initialization.

K means → Randomly // → is many cases.

↓

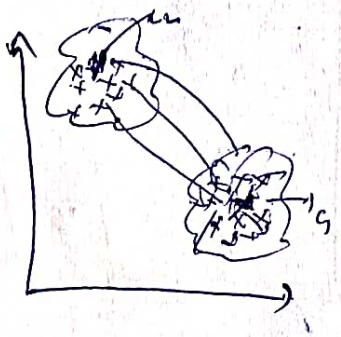
K means ++.

$C \xrightarrow{d} x_i$

Easy.

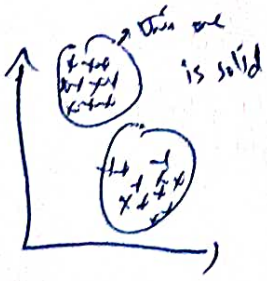
Cluster :- (1) Intra cluster distance (inside same cluster)

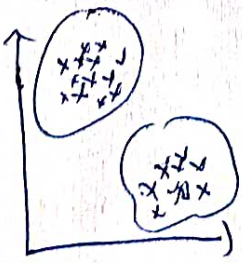
(2) Inter " " (Between a cluster)



Dunn Index :- $\frac{\max d(i,j)}{\max d(k)}$

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





↑ as much as possible ↑ max (Inter)

$=$

↓ as low as possible ↓ max (intra)

my best dunn Index :- $= \frac{\max(\text{Inter})}{\max(\text{Intra})}$

$[0, \infty]$