Lec-30 23/03/21 → Hierarchical clustering approach
- agglomerative
- divisive

short notes → Clustering validation measures

Introdn of hierarchical clustering :

→ A typical clustering analysis via partitioning data set sequentially.

→ construct nested partitions layer by layer via grouping objects into a tree of clusters, without the need to know the no. of clusters in advance.

→ The generalized distance matrix is used as clustering criteria.

2 strategies → Agglomerative (numerical)
→ divisive

i) agglomerative :

→ a bottom-up strategy.

→ initially each data object is in its own atomic cluster.

→ then merge these atomic clusters into larger and larger clusters.

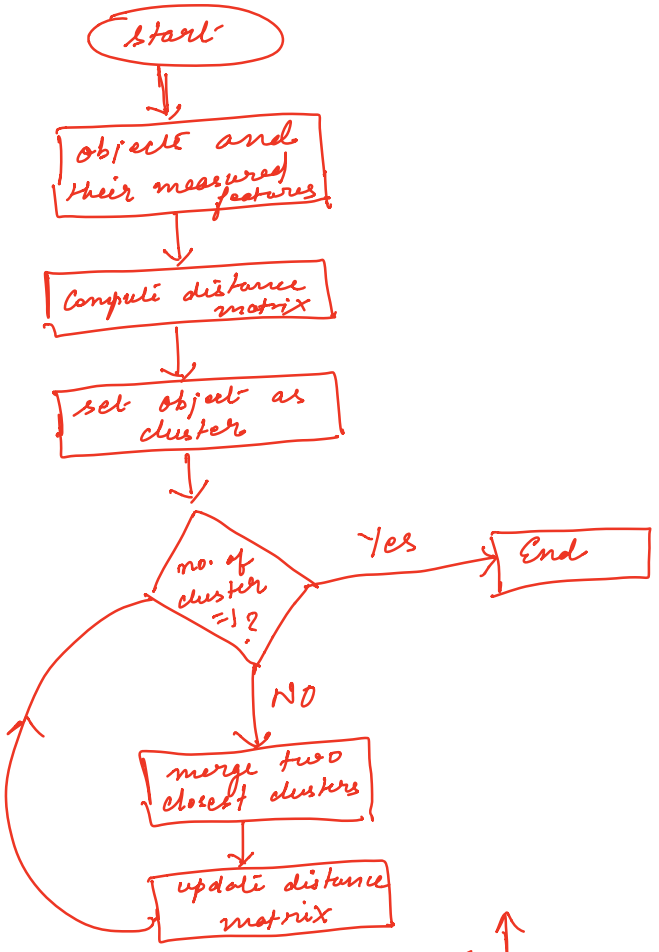
ii) divisive :

→ a top down strategy.

→ initially all objects are in one single cluster.

→ then the cluster is subdivided into smaller and smaller clusters.

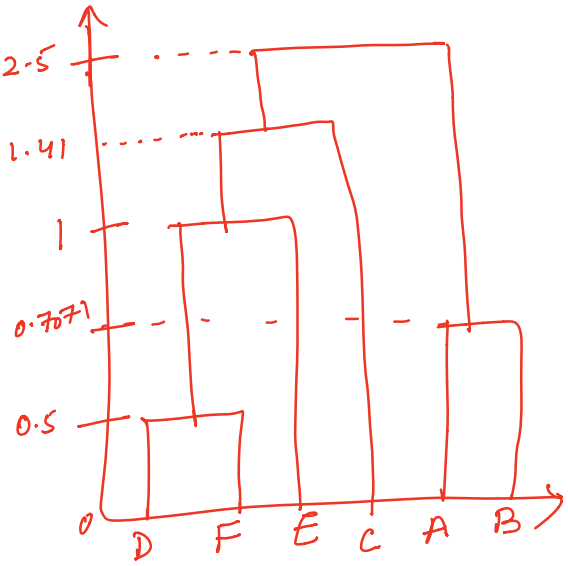
flowchart of agglomerative clustering :



Numerical :

	x_1	x_2
A	1	1
B	1.5	1.5
C	5	5
D	3	4
E	4	4
F	3	3.5

6 x 2



Item 1:

	A	B	C	D	E	F
A	0					
B	0.7071	0				
C	5.66	4.95	0			
D	3.61	2.92	2.24	0		
E	4.24	3.54	1.41	1	0	
F	3.2	2.5	2.5	0.5	1.12	0

A → B = B to A

(D and F are merged)

$$d_{AB} = \sqrt{(1-1.5)^2 + (1-1.5)^2} = 0.7071$$

$d_{CA} = 5.66$ $d_{CB} = 4.95$ $d_{DA} = 3.61$ $d_{DB} = 2.92$

$d_{DC} = 2.24$ $d_{EA} = 4.24$ $d_{EB} = 3.54$ $d_{EC} = 1.41$

$d_{ED} = 1$ $d_{FA} = 3.2$ $d_{FB} = 2.5$ $d_{FC} = 2.5$

$d_{FD} = 0.5$ $d_{FE} = 1.12$

Item 02 :

	A	B	C	DF	E
A	0				
B	0.7071	0			
C	5.66	4.95	0		
DF	3.2	2.5	2.24	0	
E	4.24	3.54	1.41	01	0

[A and B are merged]

$d_{DF \to A} = \min(DA, FA) = \min(3.61, 3.2) = 3.2$

$d_{DF \to B} = \min(DB, FB) = \min(2.92, 2.5) = 2.5$

$d_{DF \to C} = \min(DC, FC) = \min(2.24, 2.5) = 2.24$

$d_{DF \to E} = \min(DE, FE) = \min(1, 1.12) = 01$

Item 03 :

	AB	C	DF	E
AB	0			
C	4.95	0		
DF	2.5	2.24	0	
E	3.54	1.41	01	0

[DF and E are merged]

$d_{C \to AB} = \min(CA, CB) = \min(5.66, 4.95) = 4.95$

$d_{DF \to AB} = \min(DA, DB, FA, FB)$

$= \min(3.61, 2.92, 3.2, 2.5) = 2.5$

$d_{DF \to C} = 2.24$

$d_{E \to AB} = \min(EA, EB) = \min(4.24, 3.54)$

$= 3.54$

Item 04 :

	AB	C	[(DF), E]
AB	0		
C	4.95	0	
[(DF), E]	2.5	1.41	0

([(DF), E] merged with C)

Item 05 :

	AB	[[(DF), E], C]
AB	0	
[[(DF), E], C]	2.5	0