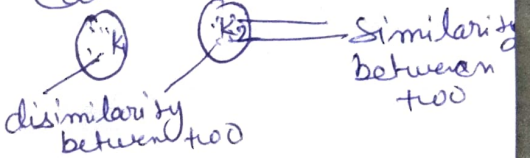
Hierarchical clustering →

Clustering → Clustering is a technique that groups Similar objects such that the objects in the same group are more Similar to Each other than the object in the other group. The Group of Similar object is called a cluster.



Hierarchical clustering Algorithm → it is a unsupervised clustering algorithm which involves creating clusters that have ~~independent~~ Predominant ordering from top to bottom. This clustering technique is divided in two types

(i) Agglomerative Hierarchical clustering

(ii) Divisive Hierarchical clustering.

(i) Agglomerative Hierarchical clustering.

It is a bottom up approach Each Observation starts in its own cluster, and Pairs of clusters are merged as one moves up the hierarchy.

two types →

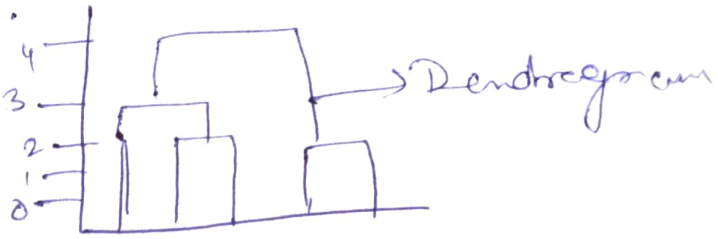
(i) Complete linkage

the distance between two clusters is defined as the longest distance between two Points in Each cluster.

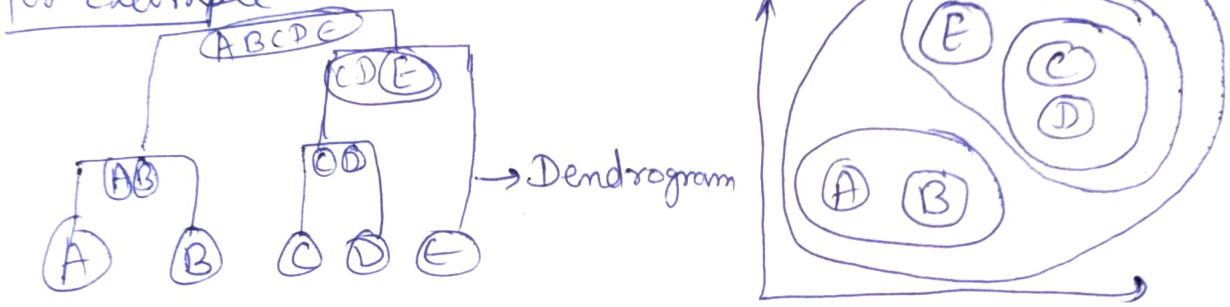
(ii) Single linkage

The distance between two clusters is defind as the shortest distance between two Points two Points in Each cluster

Dendrogram → A Dendrogram is a type of tree diagram showing hierarchical relationship between different sets of data.



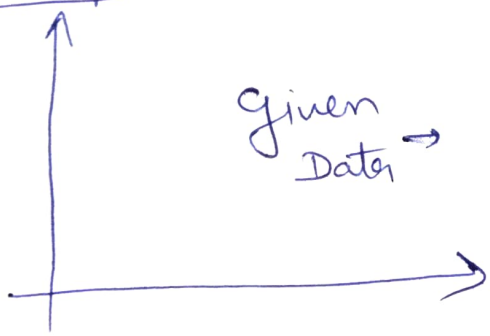
for Example



Bottom up approach

(i) **Single linkage technique →**

Example



	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	8	0

distance Matrix

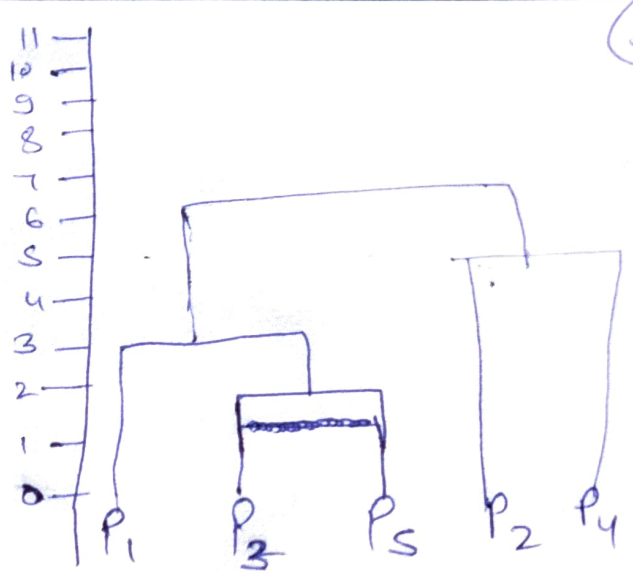
(i) We have to choose Min value in distance Matrix.

2 is Min.

2 is distance between P_3 & P_5

(ii) we have to form a cluster between P_3 & P_5 and Draw Dendrogram between P_3 & P_3

	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



* $d(P_1 [P_3 P_5])$

$\Rightarrow \min(d(P_1 P_3), d(P_1 P_5))$

$\Rightarrow \min(3, 11) \Rightarrow 3$

* $d(P_2 [P_3 P_5])$

$\Rightarrow \min(d(P_2, P_3), d(P_2, P_5))$

$\Rightarrow \min(7, 10) \Rightarrow 7$

* $d(P_4 [P_3 P_5])$

$\Rightarrow \min(d(P_4, P_3)\ d(P_4, P_5))$

$\Rightarrow \min(9, 8) \Rightarrow 8$

Now again we have to choose min value in distance matrix.

Min value — (3)

$[P_3 P_5] - [P_1] \rightarrow [P_3 P_5 P_1]$

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

* $d(P_2 [P_1 P_3 P_5])$

$\Rightarrow \min(d(P_2, P_1), d(P_2, P_3), d(P_2, P_5))$

$\Rightarrow \min(9, 7, 10)$

$\Rightarrow 7$

* $d(P_4 [P_1 P_3 P_5])$

$\Rightarrow \min(d(P_4, P_1), d(P_4, P_3), d(P_4, P_5))$

$\Rightarrow \min(6, 9, 8) \Rightarrow 6$

$\Rightarrow 6$

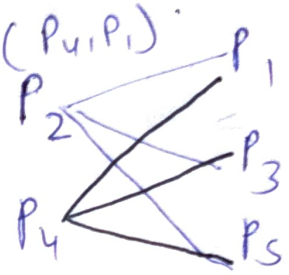
Now again we have to choose min

$P_2 - P_4 \xrightarrow{5} [P_2, P_4]$

	$[P_1 P_3 P_5]$	$[P_2 P_4]$
$[P_1 P_3 P_5]$	0	
$[P_2 P_4]$	6	0

* $d([P_1 P_3 P_5]\ [P_2 P_4])$

$\Rightarrow \min(d(P_2,P_1), d(P_2,P_3), d(P_2,P_5), d(P_4,P_1), d(P_4,P_3), d(P_4,P_5))$



$\Rightarrow \min(9, 7, 10, 6, 9, 8)$

$\Rightarrow 6$

(11) Complete linkage technique

	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	8	0

(1) we have chose min value from distance Matrix. [Except diagonal value]

2 is Min.

P_3 & P_5 we have to form cluster between P_3, P_5

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

$\Rightarrow \max(d(P_2,P_3), d(P_2,P_5))$

$\rightarrow \max(7, 10) \Rightarrow 10$

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

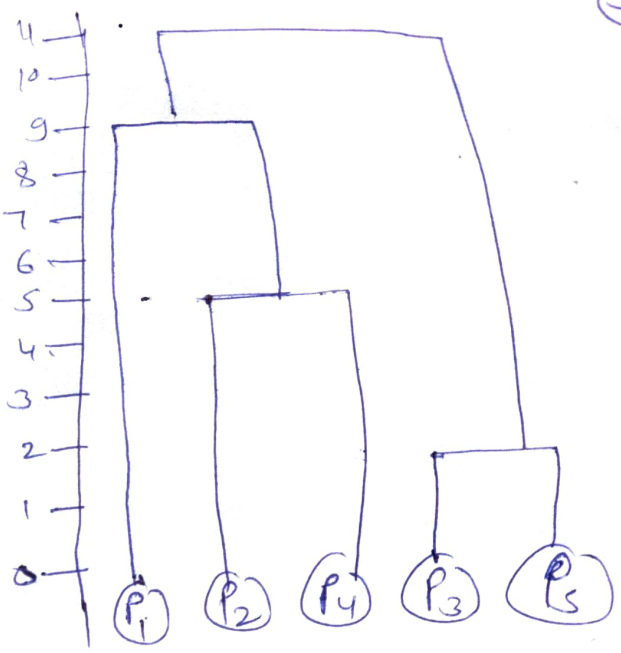
$\max(d(P_1, P_3), d(P_1, P_5))$

$\max(3, 11) \Rightarrow 11$

* $d(P_4 [P_3, P_5])$

$\max(d(P_4, P_3)\ d(P_4, P_5))$

$\max(9, 8) \Rightarrow 9$



	P_1	P_2	$[P_3 P_5]$	P_4
P_1	0			
P_2	9	0		
$[P_3 P_5]$	11	10	0	
P_4	6	(5)	9	0

Now again we have to choose min from new distance matrix $5 \rightarrow P_2\ P_4 \rightarrow [P_2, P_4]$

	P_1	$[P_2, P_4]$	$[P_3 P_5]$
P_1	0		
$[P_2 P_4]$	(9)	0	
$[P_3 P_5]$	11	10	0

* $d(P_1, [P_2, P_4])$

$\max(d(P_1, P_2), d(P_1, P_4))$

$\max(9, 6)$

$\Rightarrow \underline{9}$

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

$\max(d(P_1, P_3), d(P_1, P_5))$

$\max(3, 11) \Rightarrow \underline{11}$

* $d([P_3, P_5], [P_2 P_4])$

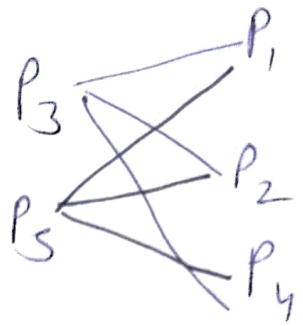
$\max((d P_3, P_2), d(P_3, P_4), d(P_5, P_2), d(P_5, P_4))$

$\max(7, 9, 10, 8)$

$\Rightarrow \underline{10}$

again we have to choose min value from current distance matrix → 9 (P1) [P2 P4] → (P1 P2 P4)

	(P1 P2 P4)	(P3 P5)
(P1 P2 P4)	0	
(P3 P5)	11	0



d ([P1 P2 P4] [P3 P5])

max (d(P3, P1), d(P3, P2), d(P3, P4), d(P5, P1), d(P5, P2), d(P5, P4))

max (3, 7, 9, 11, 10, 8)

⇒ 11

Divisive Hierarchical clustering →

* In this data object are grouped in a top down manner.
* Initially all object are in one cluster.
* Then the cluster is subdivided into smaller and smaller pieces, until each object form a cluster on its own or until it satisfies the termination condition as the desired number of clusters is obtained.

Divisive Algorithm → Simple based on the MST

1) Compute a Minimum Spanning Tree (MST) for the given adjacency Matrix.
2) Repeat-
3) Create a new cluster by breaking the link corresponding to the largest distance.

4) until only Singleton cluster remain.

Example

	A	B	C	D	E
A	0	1	2	2	3
B	1	0	2	4	3
C	2	2	0	1	5
D	2	4	1	0	3
E	3	3	5	3	0

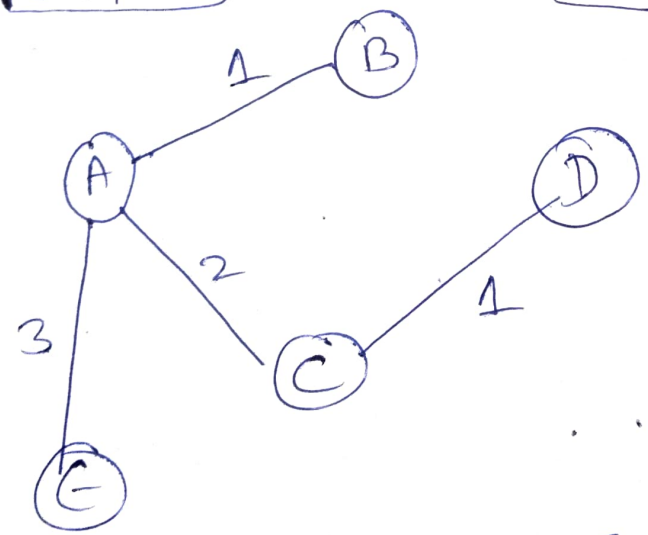
from Adjacency Matrix, Create Minimum Spanning Tree by Prim's or Kruskal's Algorithm.

by using Kruskal's

Edge	Cost
A-B	1
C-D	1
A-C	2
A-D	2
B-C	2
A-E	3
B-E	3
D-E	3
B-D	4
C-E	5

Accending order →

Edge	Cost	
A-B	1	
C-D	1	
A-C	2	
A-D	2	X
B-C	2	X
A-E	3	
B-E	3	X
D-E	3	X
B-D	4	X — forming Circle
C-E	5	X

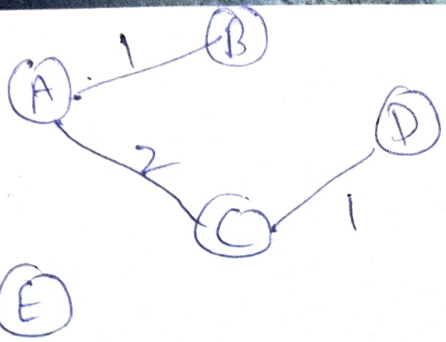


[Minimum Spanning Tree]

(i) largest Edge is between A & E

Cutting this Edge results into two clusters

{E} and {A, B, C, D}



Next remove the edge between A & C

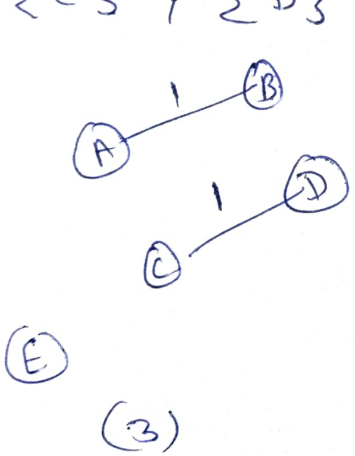
This split creates three clusters {A, B}, {C, D}, {E}

Next break A & B

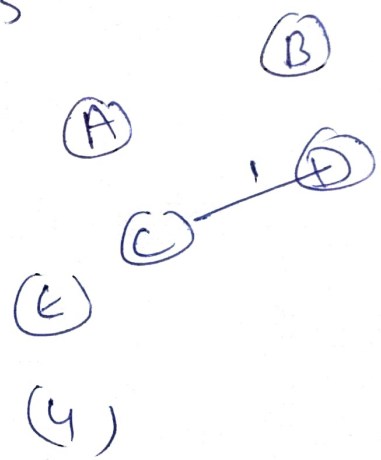
This split creates four clusters {A}, {B}, {C, D}, {E}.

Next break C & D

This split creates five clusters {A}, {B}, {C}, {D}, {E}.



(3)



(4)

(5)

Example →

1	0				
2	9	0			
3	3	7	0		
4	6	5	9	0	
5	11	10	2	8	0
	1	2	3	4	5