Agglomerative Hierarchical Clustering

Q. Assume the dataset D is given ~~by~~. Use single link technique to find the clusters in D. Use Euclidean Distance Measure.

D =

	X	Y
P_1	0.40	0.53
P_2	0.22	0.38
P_3	0.35	0.32
P_4	0.26	0.19
P_5	0.08	0.41
P_6	0.45	0.30

Step 1: Calculate the distance from each data object to all other points & create a distance Matrix.

- Euclidean Distance between points (X, Y) is given as follows.

$$d(X, Y) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Calculating ED between point P_1 & P_2 as follows.

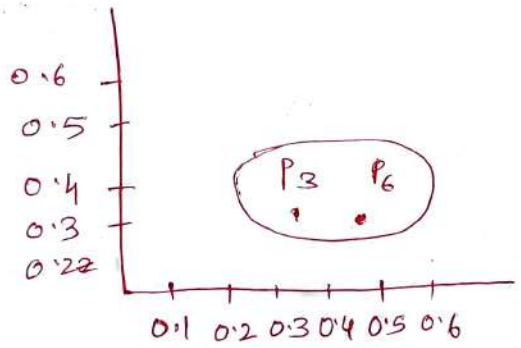
$$d(P_1, P_2) = \sqrt{(0.4 - 0.22)^2 + (0.53 - 0.38)^2}$$

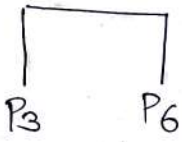
$$= 0.23$$

calculate the distance to the remaining points

	P_1	P_2	P_3	P_4	P_5	P_6
P_1	0					
P_2	0.24	0				
P_3	0.22	0.15	0			
P_4	0.37	0.20	0.15	0		
P_5	0.34	0.14	0.28	0.29	0	
P_6	0.23	0.25	0.11	0.22	0.39	0

We can note that points P_3 & P_6 are close, as the distance is minimum. so we will form a cluster with these two points.





Step 2

Recalculating the distance Matrix.

	P_1	P_2	(P_3, P_6)	P_4	P_5	P_6
P_1	0					
P_2	0.24	0				
(P_6, P_3)	0.22	0.15	0			
P_4	0.37	0.20	0.15	0		
P_5	0.34	0.14	0.28	0.29	0	
P_6						

To calculate the distance of P_1 from (P_3, P_6)

$dist((P_3, P_6), P_1) =$

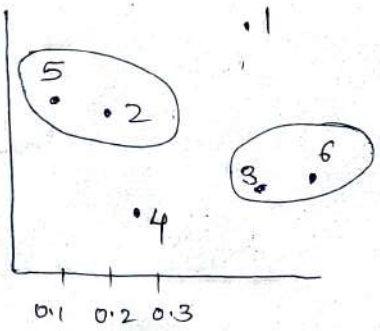
$= MIN(dist(P_3, P_1), dist(P_6, P_1))$

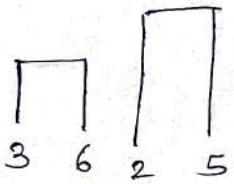
$= MIN(0.22, 0.23)$

Similarly calculate distance for all the remaining points.

Step 3: Find the next closest point & merge the same

We can see from the distance matrix that points P_2 & P_5 are close to each other.





Merging the points to form a cluster.

	P_1	P_2, P_5	P_3, P_6	P_4
P_1	0			
P_2, P_5	0.24	0		
P_3, P_6	0.22	0.15	0	
P_4	0.37	0.20	0.15	0

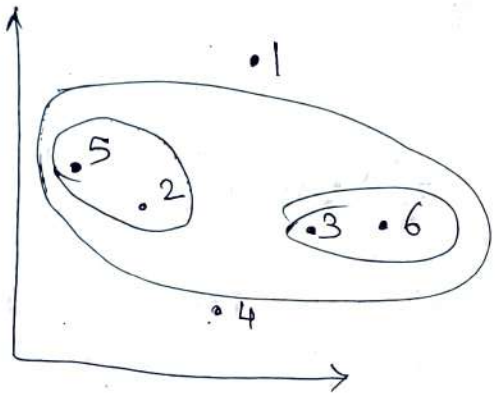
The distance between (P_3, P_6) & (P_2, P_5) is calculated as given.

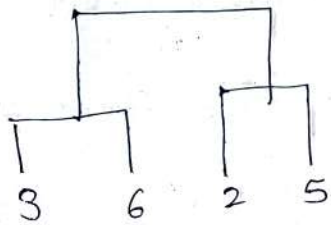
$$dist((P_3, P_6), (P_2, P_5)) = MIN(dist(P_3, P_2), dis(P_3, P_5), dist(P_6, P_2), dist(P_6, P_5))$$

$$= MIN(0.15, 0.25, 0.28, 0.39)$$

$$= 0.15$$

As can be seen from the distance matrix, $(P_2, P_5)(P_3, P_6)$ can be merged as they have minimum distance. ie 0.15

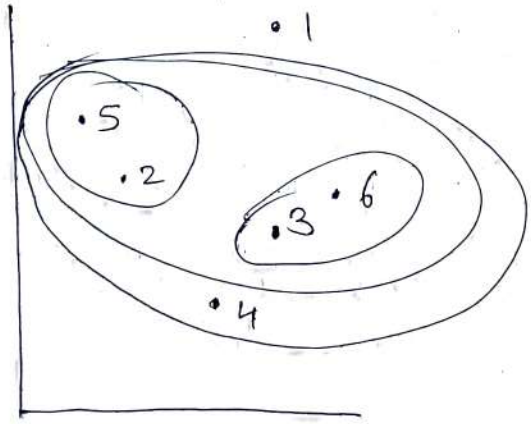




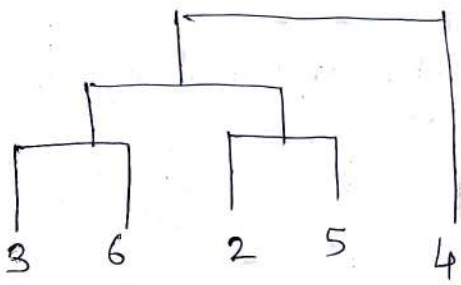
Recalculating

	P_1	(P_2, P_5, P_3, P_6)	P_4
P_1	0		
P_2, P_5, P_3, P_6	0.22	0	
P_4	0.37	0.15	0

Distance between (P_2, P_5, P_3, P_6) & P_4 is least, hence we will combine/merge these points to form the cluster.



Dendrogram



A P SHAH INSTITUTE OF TECHNOLOGY

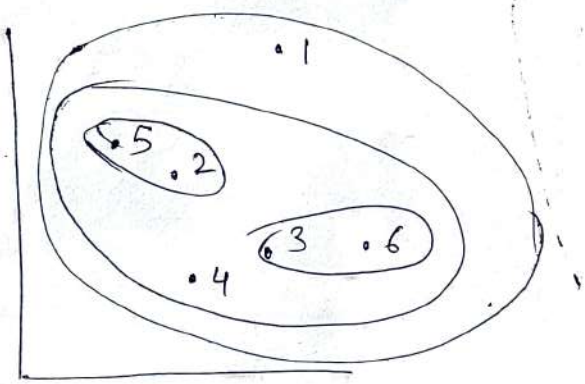
(Approved by AICTE New Delhi & Govt. of Maharashtra, Affiliated to University of Mumbai)
(Religious Jain Minority)

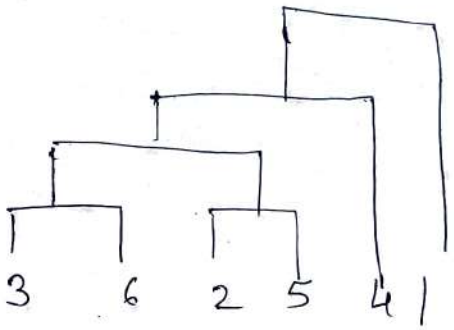
P_1 $(P_2, P_5, P_3, P_6, P_4)$

P_1

$((P_2, P_5), (P_3, P_6)), P_4)$

0	
0.22	0





[redacted]