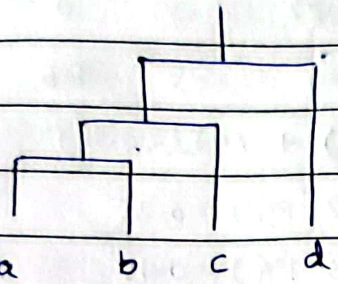
Date
Page

DWM Module 3 - Continue.

Hierarchical clustering

Agglomerative



dendogram
{a, b}
{{a, b}, c}
{{{a, b}, c}, d}

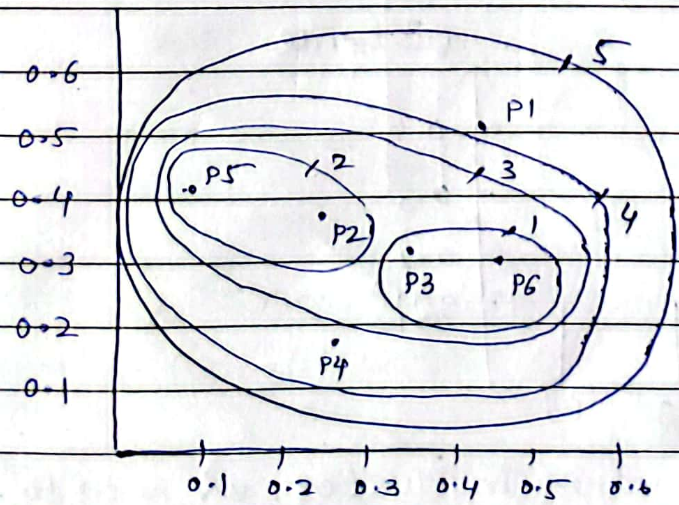
Tree Structure.

1) Single Link : Smallest distance between an element in one cluster and an element in the other.

2) Complete Link :

3) Average Link :

Ex:



points	X	Y
P1	0.40	0.53
P2	0.22	0.38
P3	0.35	0.32
P4	0.26	0.19
P5	0.08	0.41
P6	0.45	0.30

Date
Page

P1	0					
P2	0.24	0				
P3	0.22	0.15	0			
P4	0.37	0.20	0.15	0		
P5	0.34	0.14	0.28	0.29	0	
P6	0.23	0.25	0.11	0.22	0.39	0
	P1	P2	P3	P4	P5	P6

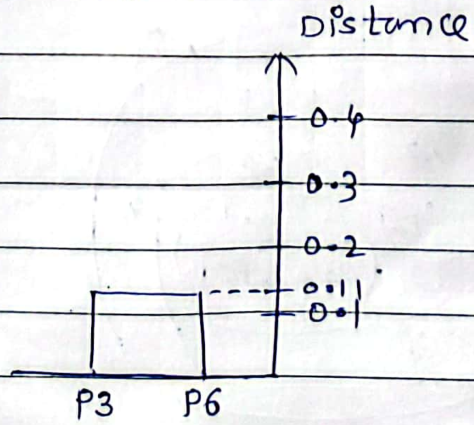
P1 = 0.40, 0.53
P2 = 0.22, 0.38

$d(P1, P2) = \sqrt{(0.40-0.22)^2 + (0.53-0.38)^2} = 0.234$

$d(P1, P2) = 0.24$ | $d(P2, P6) = 0.25$
$d(P1, P3) = 0.22$ | $d(P3, P4) = 0.15$
$d(P1, P4) = 0.37$ | $d(P3, P5) = 0.28$
$d(P1, P5) = 0.34$ | $d(P3, P6) = 0.11$
$d(P1, P6) = 0.23$ | $d(P4, P5) = 0.29$
$d(P2, P3) = 0.15$ | $d(P4, P6) = 0.22$
$d(P2, P4) = 0.20$ | $d(P5, P6) = 0.39$
$d(P2, P5) = 0.14$

Identify the clusters with the smallest distance in matrix and merge them together

(I)



P3, P6 merge together in cluster, we need to recompute the distance from each point to our new cluster (P3, P6)

	P1	P2	P3, P6	P4	P5
P1	0				
P2	0.24	0			
P3, P6	0.22	0.15	0		
P4	0.37	0.20	0.15	0	
P5	0.34	0.14	0.28	0.29	0

$$dist((P3,P6), P1) = min(dist(P3,P1), dist(P6,P1))$$
$$= min(0.22, 0.23)$$
$$= \underline{0.22}$$

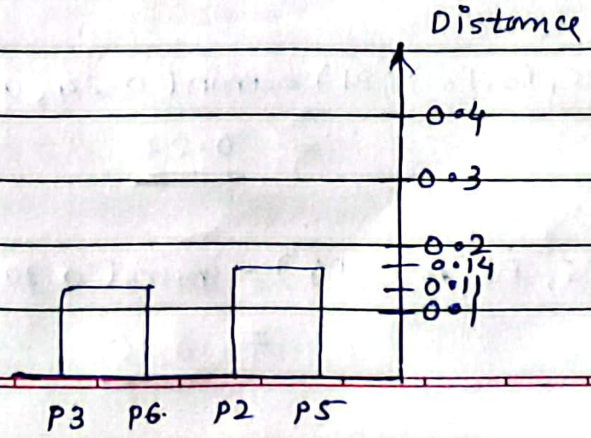
$$dist((P3,P6), P2) = min(dist(P3,P2), dist(P6,P2))$$
$$= min(0.15, 0.25)$$
$$= \underline{0.15}$$

$$dist((P3,P6), P4) = min(dist(P3,P4), dist(P6,P4))$$
$$= min(0.15, 0.22)$$
$$= \underline{0.15}$$

$$dist((P3,P6), P5) = min(dist(P3,P5), dist(P6,P5))$$
$$= min(0.28, 0.39)$$
$$= \underline{0.28}$$

Now, (P2, P5) have smallest distance so merge those two in a single cluster.

(II)



P1	0			
P2,P5	0.24	0		
P3,P6	0.22	0.15	0	
P4	0.37	0.20	0.15	0
	P1	P2,P5	P3,P6	P4

dist ((P2,P5),(P3,P6)) = min (d(P2,P3), d(P2,P6), d(P5,P3), d(P5,P6))

= min (0.15, 0.25, 0.28, 0.39)

= 0.15

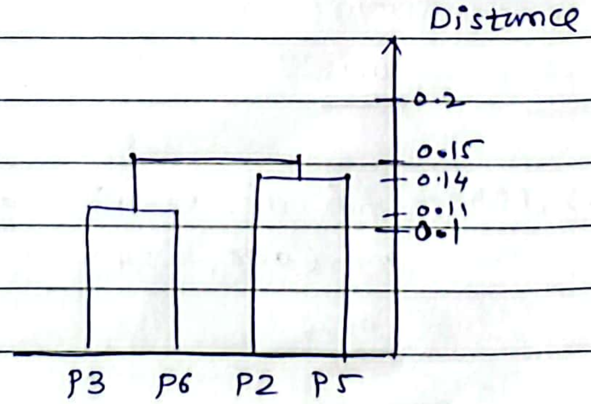
dist ((P2,P5), P4) = min (d(P2,P4), d(P5,P4))

= min (0.20, 0.29)

= 0.20

dist ((P2,P5), P1) = min (0.24, 0.34) = 0.24

(III)



Now, recompute the distance from each point to new cluster (P2, P5, P3, P6)

dist ((P2,P5, P3,P6), P1) = min (0.24, 0.34, 0.22, 0.23)

= 0.22

dist ((P2,P5, P3,P6), P4) = min (0.20, 0.29, 0.15, 0.22)

= 0.15

(IV)

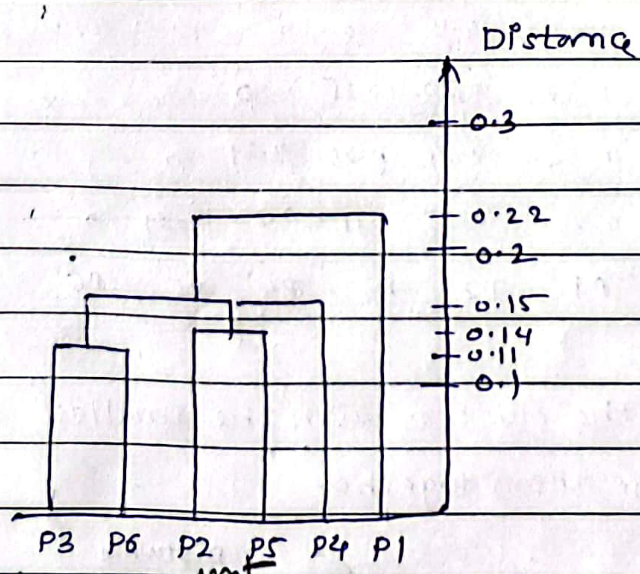
P1	0		
P2, P3, P5, P6	0.22	0	
P4	0.37	0.15	0
	P1	P2, P3, P5, P6	P4

Now, recompute the distance from each point to new cluster (P2, P5, P3, P6, P4)

$$dist((P2, P5, P3, P6, P4), P1) = \min(0.24, 0.34, 0.22, 0.23, 0.37)$$
$$= \underline{\underline{0.22}}$$

P1	0	
P2, P5, P3, P6, P4	0.22	0
	P1	P2, P5, P3, P6, P4

(V)



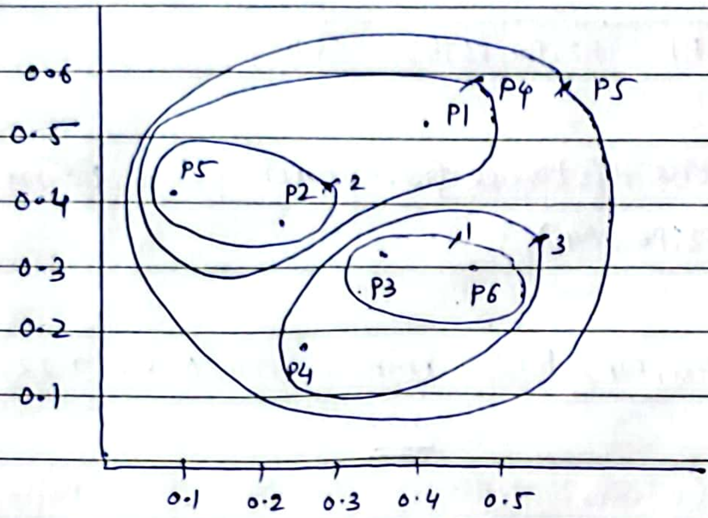
clusters ↓ ↙ independent points

DE = { <0, 6 {{P1}, {P2}, {P3}, {P4}, {P5}, {P6} }>,
<1, 5 {{P1}, {P2}, {P3, P6}, {P4}, {P5}}>,
<2, 4 {{P1}, {P2, P5}, {P3, P6}, {P4} }>,
<3, 3 {{P1}, {P2, P5, P3, P6}, {P4}}>,
<4, 2 {{P1}, {P2, P5, P3, P6, P4}}>,
<5, 1 {{P1, P2, P5, P3, P6, P4}}> }

2) Complete Link : Largest distance between an element in one cluster and an element in other.

(same as previous Question)

Ex.

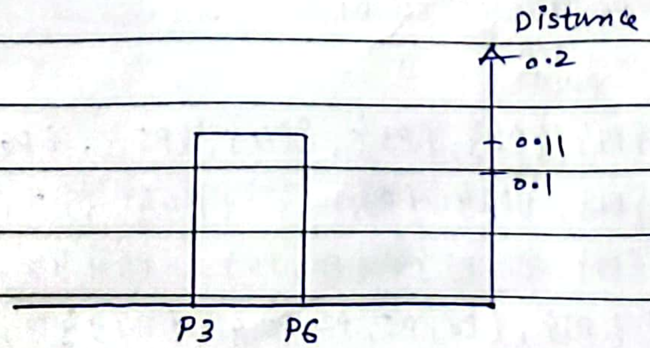


points	X	Y
P1	0.40	0.53
P2	0.22	0.38
P3	0.35	0.32
P4	0.26	0.19
P5	0.08	0.41
P6	0.45	0.30

	P1	P2	P3	P4	P5	P6
P1	0					
P2	0.24	0				
P3	0.22	0.15	0			
P4	0.37	0.20	0.15	0		
P5	0.34	0.14	0.28	0.29	0	
P6	0.23	0.25	0.11	0.22	0.39	0

Identify the clusters with the smallest distance in matrix and merge them together

(I)



P3, P6 merge together in cluster, we need to recompute the distance from each point to our new cluster (P3, P6)

	P1	P2	P3,P6	P4	P5
P1	0				
P2	0.24	0			
P3,P6	0.23	0.25	0		
P4	0.37	0.20	0.22	0	
P5	0.34	0.14	0.39	0.29	0

$$\text{dist}((P3,P6),P1) = \max(d(P3,P1), d(P6,P1))$$
$$= \max(0.22, 0.23)$$
$$= \underline{0.23}$$

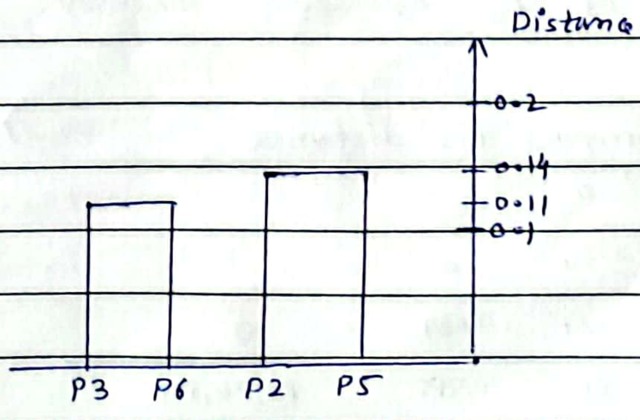
$$\text{dist}((P3,P6),P2) = \max(d(P3,P2), d(P6,P2))$$
$$= \max(0.15, 0.25)$$
$$= \underline{0.25}$$

$$\text{dist}((P3,P6),P4) = \max(d(P3,P4), d(P6,P4))$$
$$= \max(0.15, 0.22)$$
$$= \underline{0.22}$$

$$\text{dist}((P3,P6),P5) = \max(d(P3,P5), d(P6,P5))$$
$$= \max(0.28, 0.39)$$
$$= \underline{0.39}$$

Now (P2,P5) have smallest distance so merge those two in single cluster.

(II)



P1	0			
P2, P5	0.34	0		
P3, P6	0.23	0.39	0	
P4	0.37	0.29	0.22	0
	P1	P2, P5	P3, P6	P4

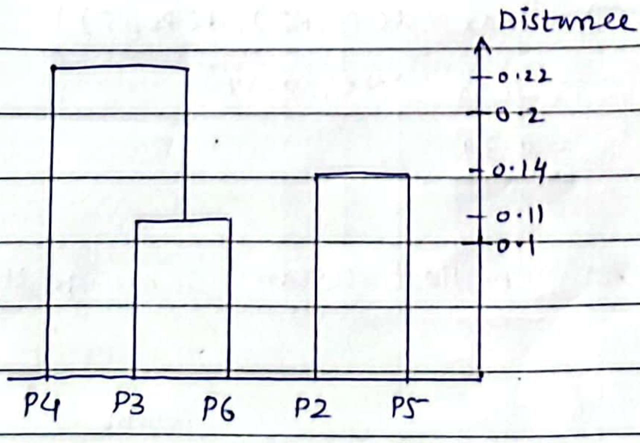
$$\text{dist}((P2, P5), P1) = \max(0.24, 0.34) = \underline{0.34}$$

$$\text{dist}((P2, P5), (P3, P6)) = \max(0.15, 0.25, 0.28, 0.39) = \underline{0.39}$$

$$\text{dist}((P2, P5), P4) = \max(0.20, 0.29) = \underline{0.29}$$

Now ((P3, P6), P4) have smallest distance so merge those two in a single cluster

(III)



Now recompute the distance

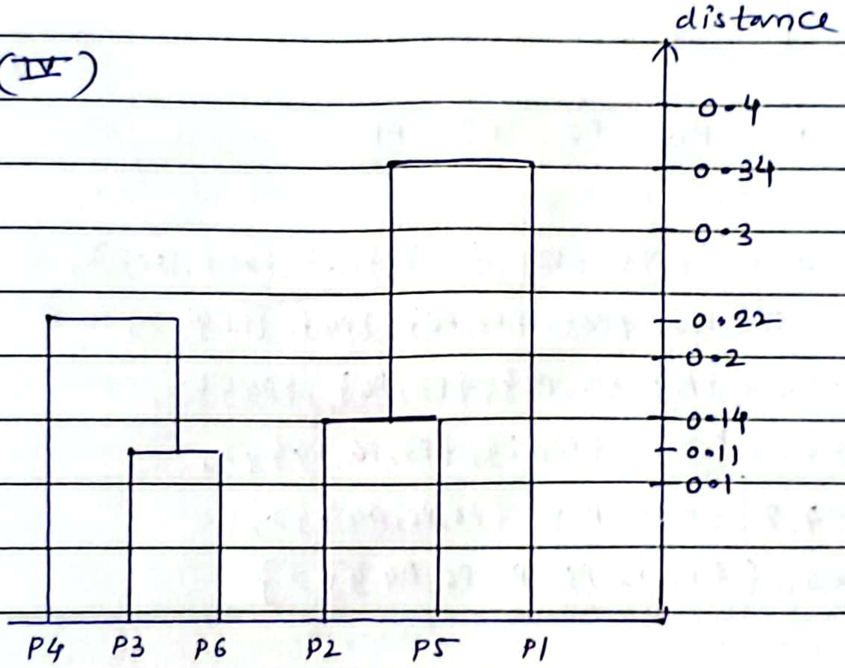
P1	0		
P2, P5	0.34	0	
P3, P6, P4	0.37	0.39	0
	P1	P2, P5	P3, P6, P4

$$dist((P3,P6,P4),P1) = \max(0.22, 0.23, 0.37)$$
$$= \underline{0.37}$$

$$dist((P3,P6,P4),(P5,P2)) = \max(0.15, 0.28, 0.25, 0.39, 0.20, 0.29)$$
$$= \underline{0.39}$$

Now, ((P2,P5),P1) have smallest distance so merge those two in a single cluster.

(IV)

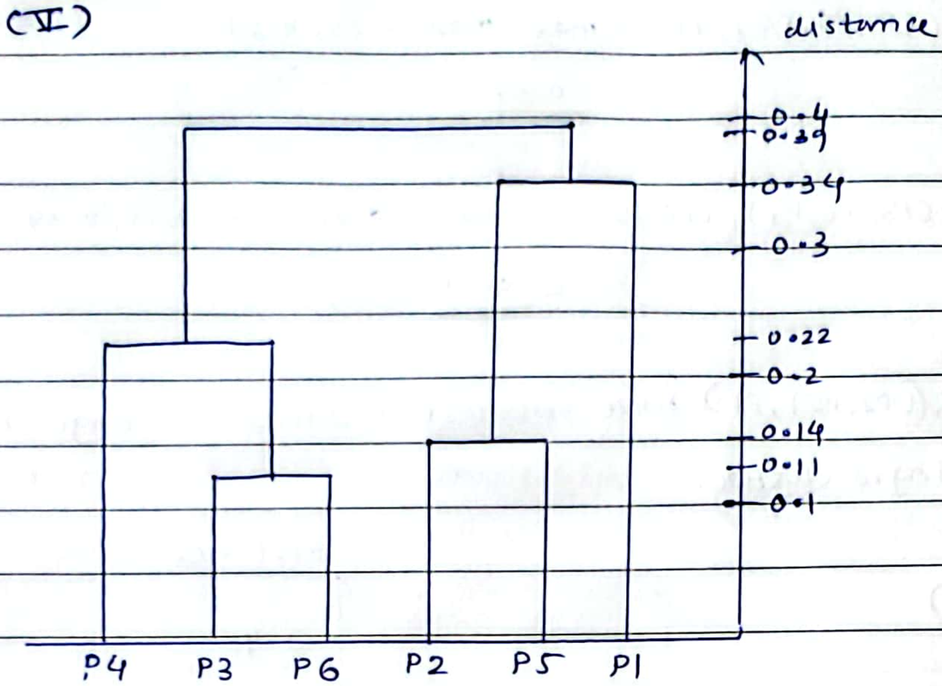


Now, recompute the distance

	P1, P2, P5	P3, P6, P4
P1, P2, P5	0	
P3, P6, P4	0.39	0

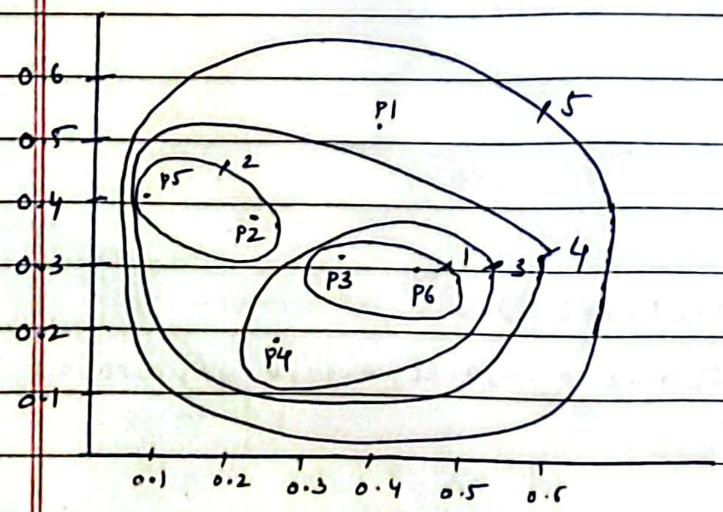
$$dist((P3,P6,P4),(P1,P2,P5))$$
$$= \max(0.22, 0.15, 0.28, 0.23, 0.25, 0.39, 0.37, 0.20, 0.29)$$
$$= \underline{0.39}$$

(II)



DE = { <0, 6 {{P1}, {P2}, {P3}, {P4}, {P5}, {P6}}>,
<1, 5 {{P1}, {P2}, {P3, P6}, {P4}, {P5}}>,
<2, 4 {{P1}, {P2, P5}, {P3, P6}, {P4}}>,
<3, 3 {{P1}, {P2, P5}, {P3, P6, P4}}>,
<4, 2 {{P1, P2, P5}, {P3, P6, P4}}>,
<5, 1 {{P1, P2, P5, P3, P6, P4}}> }

3) Average Link : Follow Average Link technique to find clusters in D.



points	X	Y
P1	0.40	0.53
P2	0.22	0.38
P3	0.35	0.32
P4	0.26	0.19
P5	0.08	0.41
P6	0.45	0.30

	P1	P2	P3	P4	P5	P6
P1	0					
P2	0.24	0				
P3	0.22	0.15	0			
P4	0.37	0.20	0.15	0		
P5	0.34	0.14	0.28	0.29	0	
P6	0.23	0.25	0.11	0.22	0.39	0

d(P1, P2) = 0.24

d(P1, P3) = 0.22

d(P1, P4) = 0.37

d(P1, P5) = 0.34

d(P1, P6) = 0.23

d(P2, P3) = 0.15

d(P2, P4) = 0.20

d(P2, P5) = 0.14

d(P2, P6) = 0.25

d(P3, P4) = 0.15

d(P3, P5) = 0.28

d(P3, P6) = 0.11

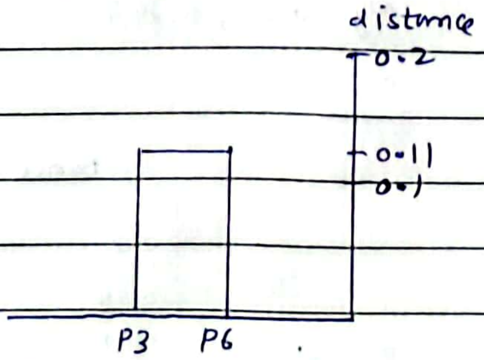
d(P4, P5) = 0.29

d(P4, P6) = 0.22

d(P5, P6) = 0.39

Identify the cluster with the smallest distance in matrix and merge them together

(I)



(P3, P6) merge together in cluster, we need to ~~ar~~ recompute the distance from each point to our new clusters (P3, P6)

P1	0				
P2	0.24	0			
P3, P6	0.23	0.20	0		
P4	0.37	0.20	0.19	0	
P5	0.34	0.14	0.34	0.29	0
	P1	P2	P3, P6	P4	P5

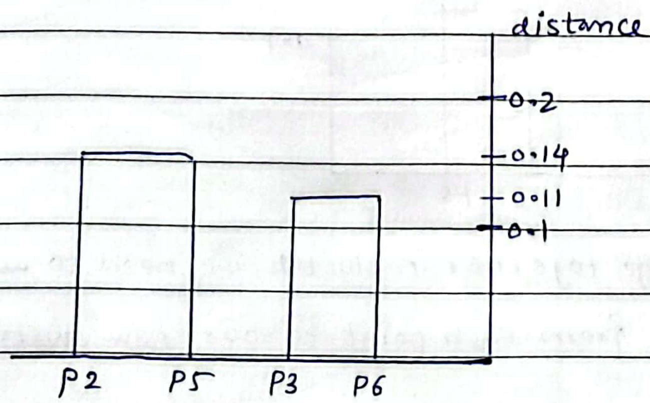
dist ((P3, P6), P1) = Avg (d(P3, P1), d(P6, P1))
= Avg (0.22, 0.23)
= 0.23

dist ((P3, P6), P2) = Avg (d(P3, P2), d(P6, P2))
= Avg (0.15, 0.25)
= 0.20

dist ((P3, P6), P4) = Avg (d(P3, P4), d(P6, P4))
= Avg (0.15, 0.22)
= 0.19

dist ((P3, P6), P5) = Avg (d(P3, P5), d(P6, P5))
= Avg (0.28, 0.39)
= 0.34

(II)



P1	0			
P2, P5	0.29	0		
P3, P6	0.23	0.27	0	
P4	0.37	0.25	0.19	0
	P1	P2, P5	P3, P6	P4

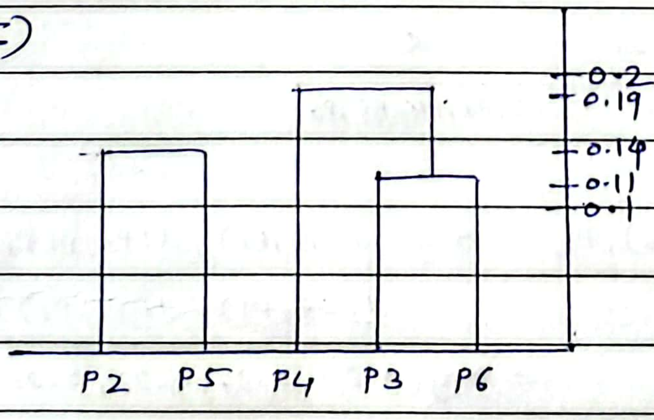
dist ((P2, P5), P1) = Avg (d(P2, P1), d(P5, P1))
= Avg (0.24, 0.34)
= 0.29

dist ((P2, P5), (P3, P6)) = Avg (d(P2, P3), d(P2, P6), d(P5, P3), d(P5, P6))
= Avg (0.15, 0.25, 0.28, 0.39)
= 0.27

dist ((P2, P5), P4) = Avg (d(P2, P4), d(P4, P5))
= Avg (0.20, 0.29)
= 0.25

As 0.19 is smallest merge((P3, P6), P4)

(III)



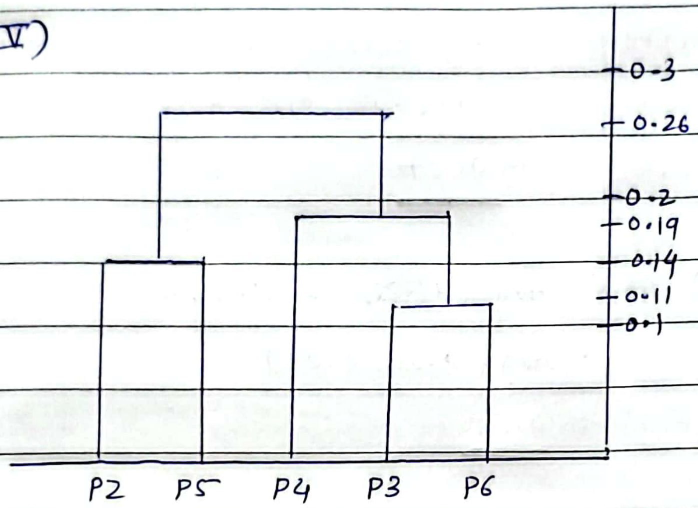
P1	0		
P2, P5	0.29	0	
P3, P6, P4	0.27	0.26	0
	P1	P2, P5	P3, P6, P4

$dist((P3,P6,P4),P1) = Avg(d(P3,P1), d(P6,P1), d(P4,P1))$
$= Avg(0.22, 0.23, 0.37)$
$= \underline{0.27}$

$dist((P3,P6,P4),(P2,P5)) = Avg(d(P3,P2), d(P6,P2), d(P4,P2),$
$d(P5,P3), d(P6,P5), d(P4,P5))$
$= Avg(0.15, 0.25, 0.20, 0.34, 0.20, 0.29)$
$= \underline{0.26}$

As 0.26 is smallest, merge (P2, P5) and (P3, P6, P4).

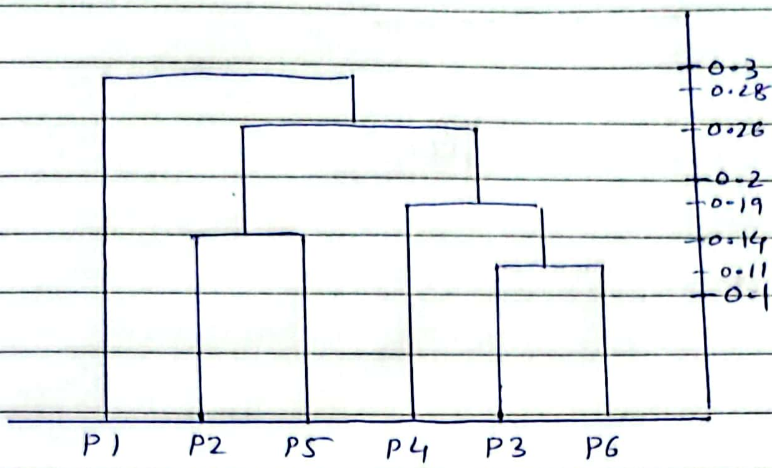
(IV)



	P1	P2,P3,P4,P5,P6
P1	0	
P2,P3,P4,P5,P6	0.28	0

$dist((P2,P3,P4,P5,P6),P1) = Avg(d(P2,P1), d(P3,P1), d(P4,P1),$
$d(P5,P1), d(P6,P1))$
$= Avg(0.24, 0.34, 0.22, 0.23, 0.37)$
$= \underline{0.28}$

(II)



DE = { <0, 6 {{P1}, {P2}, {P3}, {P4}, {P5}, {P6}}>,

<01, 5 {{P1}, {P2}, {P3, P6}, {P4}, {P5}}>,

<2, 4 {{P1}, {P2, P5}, {P3, P6}, {P4}}>,

<3, 3 {{P1}, {P2, P5}, {P3, P6, P4}}>,

<4, 2 {{P1}, {P2, P5, P3, P6, P4}}>,

<5, 1 {{P1, P2, P5, P3, P6, P4}}> }

Algorithm (Agglomerative)

Input

$D = \{t1, t2, \ldots, (tn)\}$

A

Output

DE

Algorithm

$D = 0$

$k = n$

$K = \{\{t1\}, \ldots \{tn\}\}$

$DE = \{\langle d, k, K \rangle\}$

repeat

old $k = K$;

$d = d + 1$

Ad = Vertex adjacency matrix graph with ~~wt~~ threshold distance of D.

$\langle (k, K) \rangle$ = New clusters (Ad, D);

if old $k \neq k$ then

$DE = DE \cup \langle d, k, K \rangle$;

output $= 1$.