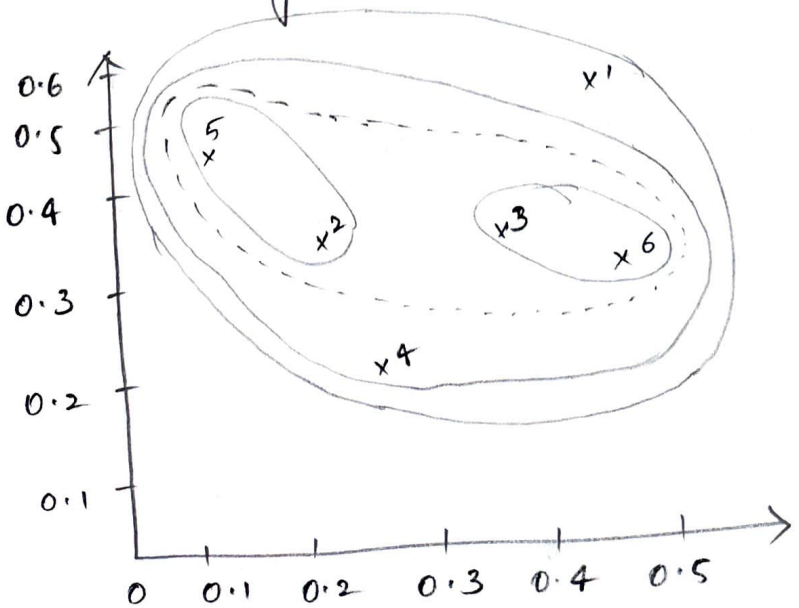
Question 1

Calculate the clustering representation & dendrogram using single, complete and average link proximity function in hierarchical clustering technique

(i) Single link proximity function

	P_1	P_2	P_3	P_4	P_5	P_6
P_1	0					
P_2	0.2357	0				
P_3	0.2218	0.1483	0			
P_4	0.3688	0.2042	0.1513	0		
P_5	0.3421	0.1388	0.2843	0.2932	0	
P_6	0.2347	0.2540	(0.1100)	0.2216	0.3921	0

Considering lowerband values, the upper bound value are equal to lowerboand & the cluster is found according to that.



Graphical representation for the given points

∴ Since we found the euclidian distance for the each points, we'll find the merging the two closest numbers & updating the distance table.

⇒ We can observe, from the above table the minimum value is 0.1100 i.e, between P_3 & P_6 after merging 2 numbers we need to find the distance with other member using the formula

$$\text{Min}[\text{dist}(\text{clusters}, \text{otherpoints})]$$

⇒ Distance with

⇒ P_1 ⇒ $\text{Min}[\text{dist}(P_3, P_6), P_1)]$
$= \text{Min}\{\text{dist}(P_3, P_1), (P_6, P_1)\}$
$= \text{Min}[0.2218, 0.2347]$
$= 0.2218$ //

⇒ P_2 ⇒ $\text{Min}[\text{dist}(P_3, P_6), P_2)]$
$= \text{Min}[\text{dist}(P_3, P_2), (P_6, P_2)]$
$= \text{Min}[0.1483, 0.2540)$
$\therefore 0.1483$ //

⇒ P_4 ⇒ $\text{Min}[\text{dist}(P_3 P_6), P_4]$
⇒ $\text{Min}[\text{dist}(P_3, P_4), (P_6, P_4)]$
⇒ $\text{Min}[0.1513, 0.2216]$
⇒ 0.1513 //

⇒ P_5 ⇒ $\text{Min}[\text{dist}(P_3, P_4), P_5]$
⇒ $\text{Min}[\text{dist}(P_3, P_5), (P_4, P_5)]$
⇒ $\text{Min}(0.2843, 0.3921)$
⇒ 0.2843 //

Therefore, the distance table after merging P_3 & P_4

	P_1	P_2	P_3P_6	P_4	P_5
P_1	0				
P_2	0.2357	0			
P_3P_6	0.2218	0.1483	0		
P_4	0.3688	0.2042	0.1513	0	
P_5	0.3421	0.1388 (Minimum)	0.2843	0.2932	0

(2) Considering the above table which is updated 0.1388 is the minimum value then join P_2 & P_5

⇒ Distance with:

P_1 ⇒ Min [dist $((P_2P_5), P_1)$]
⇒ Min $(cP_2, P_1)(P_5, P_1)$]
⇒ Min [0.2357, 0.3421]
⇒ 0.2357//

P_3P_6 ⇒ Min [dist $(P_2P_5)(P_3P_6)$]
⇒ Min [dist $((P_2P_5)(P_3P_6)$] ⇒ Min [$(P_2(P_3P_6))$ $(P_5(P_3P_6))$]
⇒ Min [0.1483, 0.2843)
⇒ 0.1483//

P_4 ⇒ Min [dist $((P_2P_5)P_4)$]
⇒ Min [dis (P_2P_4) (P_5P_4)]
⇒ (0.2042, 0.2932)
⇒ 0.2042//

So, the updated distance table after merging P_2 with P_5 is

	P_1	P_2P_5	P_3P_6	P_4
P_1	0			
P_2P_5	0.2357	0		
P_3P_6	0.2218	0.1483 (Minimum)	0	
P_4	0.3688	0.2042	0.1513	0

From the above table, since the minimum value is 0.1483, merg (P_2P_5) & (P_3P_6) and update distance matrix

⇒ Distance of $(P_2P_5)(P_3P_6)$ with

$P_1 \Rightarrow Min[dist((P_2P_5)(P_2P_6)P_1))]$

$\Rightarrow Min[dist(P_2P_5)P_1)((P_3P_6)(P_1))]$

$\Rightarrow Min[0.2357, 0.2218)$

$\Rightarrow 0.2218$ //

$P_4 \Rightarrow Min[dist((P_2P_5)(P_3P_6)P_4)]$

$\Rightarrow Min[dist((P_2P_5)P_4))]$

$\Rightarrow Min[dist((P_2P_5)P_4)((P_3P_6)P_4)))$

$\Rightarrow Min(0.2042, 0.1513)$

$\Rightarrow 0.1513$ //

Therefore, the updated distance table is

	P_1	P_2 P_5 P_3 P_6	P_4
P_1	0		
P_2 P_5 P_3 P_6	0.2218	0	
P_4	0.3688	0.1513 (minimum)	0

Since P_4 has the minimum value, merging & finding the distance from

$P_1 \Rightarrow$ Min [dist (P_2 P_5 P_3 P_6) P_4 P_1]

$\Rightarrow$ Min [dist ((P_2 P_5 P_2 P_6) P_1) (P_4 P_1)]

$\Rightarrow$ (0.2218, 0.3688)

$\rightarrow$ 0.2218

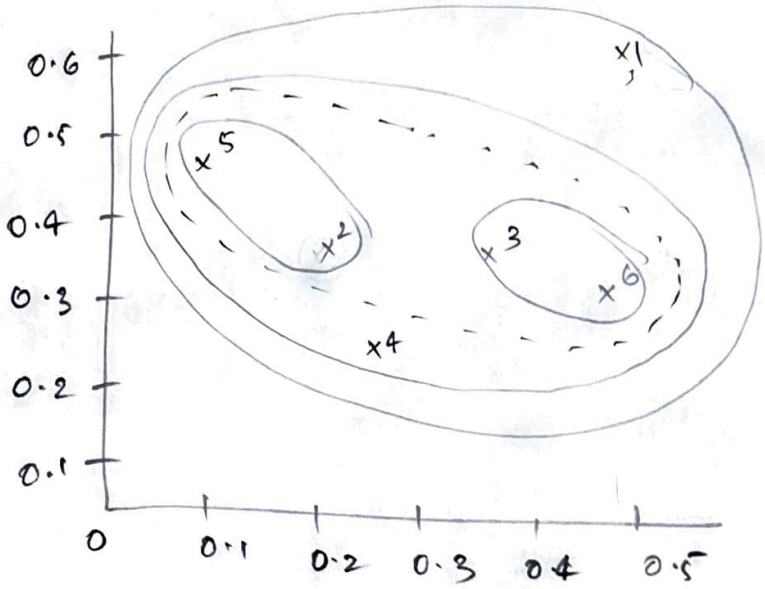
	P_1	P_2 P_5	P_3 P_6	P_4
P_1	0			
P_2 P_5 P_3 P_6 P_4	0.2218	0		

The minim distan P_1

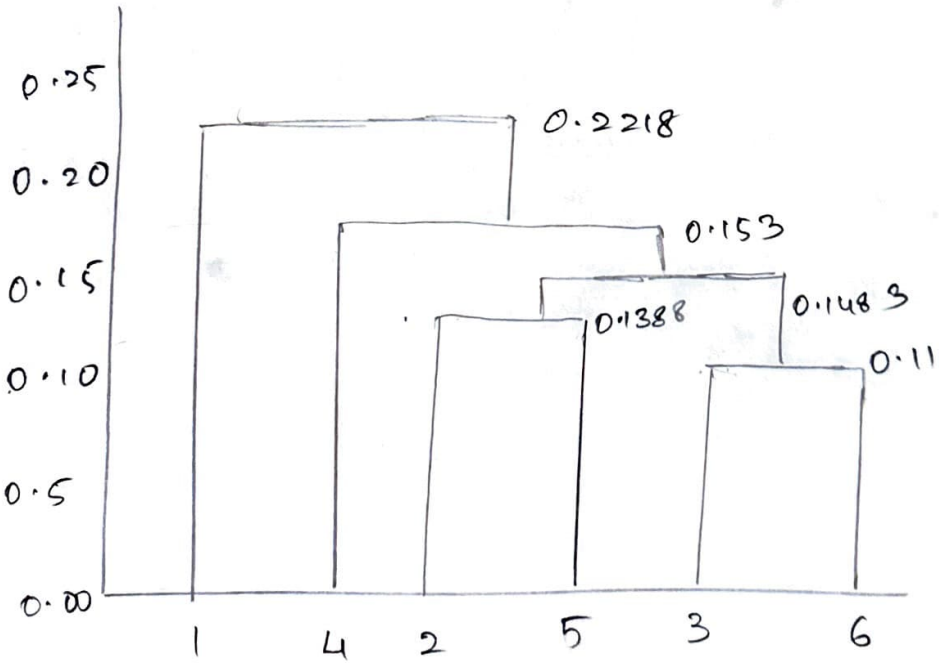
Min [dist (P_2 P_5 P_3 P_6 P_4) (P_1))

$\Rightarrow$ 0.22

∴ The final cluster is



∴ The dendogram is

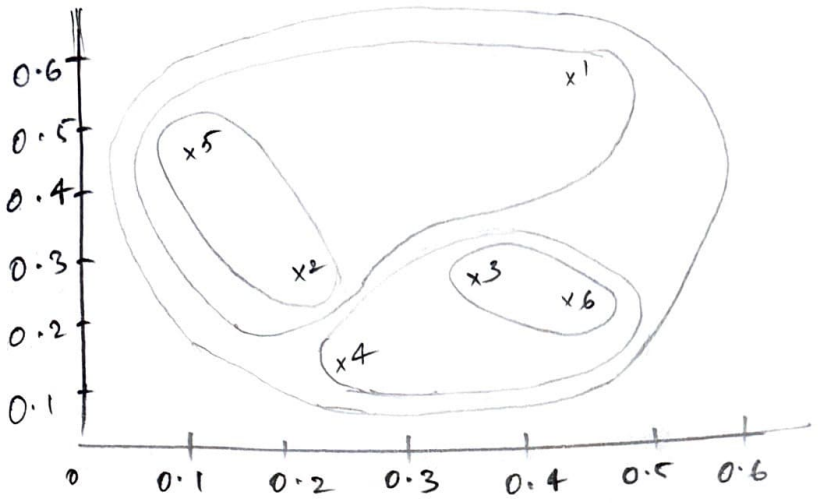


8 (ii) Complete link proximity function

For complete link proximity we need merge the closest point & update the distance table with ~~before~~ with below formula.

$$\text{Distance} \Rightarrow \text{Max}(\text{dist}(\text{clustering point}, \text{points}))$$

Graphical representation of given point:



From table above in page ① min. value is 0.1100 . i.e., merge P_3 & P_6 & update the distance table

Distance with P_1

$P_1 \Rightarrow \text{Max}[\text{dist}((P_3\ P_6), P_1)]$

$\Rightarrow \text{Max}[\text{dist}(P_3\ P_1)(P_6\ P_1)]$

$\Rightarrow \text{Max}[0.2218, 0.2347)$

$\Rightarrow 0.2347$

$P_2 \Rightarrow \text{Max}[\text{dist}(P_3\ P_6)\ P_2)]$

$\Rightarrow \text{Max}[\text{dist}((P_3 P_2)(P_6, P_2)]$

$\Rightarrow \text{Max}[0.1483, 0.2540]$

$\Rightarrow 0.2540.$

$P_4 \Rightarrow Max[dist(P_3P_6)P_4)]$

$\rightarrow Max\ [0.1513, 0.2216]$

$\Rightarrow 0.2216$ //

$P_5 \Rightarrow Max\ [dist(P_3P_6)P_5)]$

$\Rightarrow Max[0.2843, 0.342]$

$\Rightarrow 0.3421$ //

So, the distance table after the minimum distance calculations & merging P_3 & P_6 is

	P_1	P_2	P_3P_6	P_4	P_5
P_1	0				
P_2	0.2357	0			
P_3P_6	0.2347	0.2540	0		
P_4	0.3688	0.2042	0.2216	0	
P_5	0.3421	0.1388	0.3421	0	

Minimum

Minimum value is 0.1388, merge P_2P_5 & update the distance table.

Distance of (P_2P_5) with

$P_1 \Rightarrow Max[dist((P_2,P_5)P_1)]$

$\Rightarrow Max[0.2357, 0.3421]$

$\Rightarrow 0.3421$ //

$P_3P_6 \Rightarrow Max[dist((P_3P_5)(P_3P_6))]$

$\Rightarrow Max[0.2540, 0.3421]$

$\Rightarrow 0.3921$

$P_4 \Rightarrow Max[dist((P_2P_5)P_4)]$

$\Rightarrow Max[0.2042, 0.2932]$

$\Rightarrow 0.2932$ //

The distance table after merging P_2 & P_5

	P_1	P_2P_5	P_3P_6	P_4
P_1	0			
P_2P_5	0.3421	0		
P_3P_6	0.2347	0.3921	0	
P_4	0.3688	0.2932	0.2216 (Minimum)	0

Minimum value is 0.2216, merge & update the distance table.

Distance of $P_4(P_3P_6)$ with

$P_1 \Rightarrow Max\ (dist\ ((P_3P_6)P_4)\ P_1)]$

$\Rightarrow Max[0.2347, 0.3688]$

$\Rightarrow 0.3688$ //

$P_2P_5 \Rightarrow Max\ [(P_4(P_3P_6))(P_2P_5))]$

$\Rightarrow Max[0.2932, 0.3921]$

$\Rightarrow 0.3921$

So, the distance table after the minimum distance & merging.

	P_1	P_2P_5	$P_3P_4P_6$
P_1	0		
P_2P_5	0.3421 (Minimum.)	0	
$P_3P_4P_6$	0.3688	0.3921	0

Minimum value is 0.3421, so merge P_2P_5 and P_1

Distance of $P_1(P_2P_5)$ with.

$P_3P_6P_4 \rightarrow Dist[Max((P_1(P_2P_5))((P_3P_6)P_4)]$

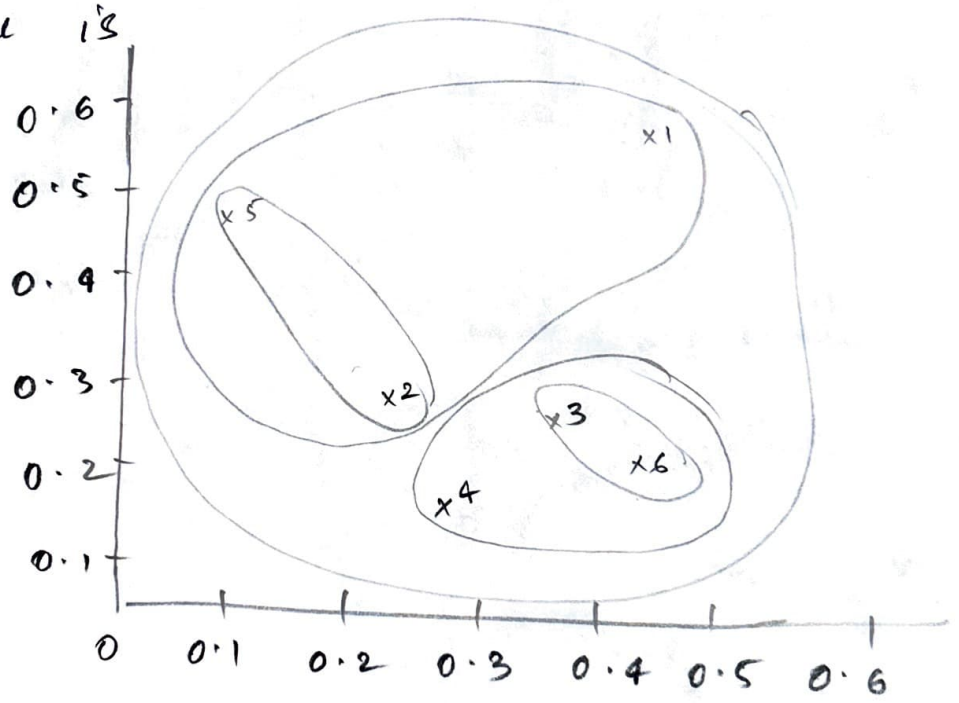
$\Rightarrow Max[0.3688, 0.3921]$

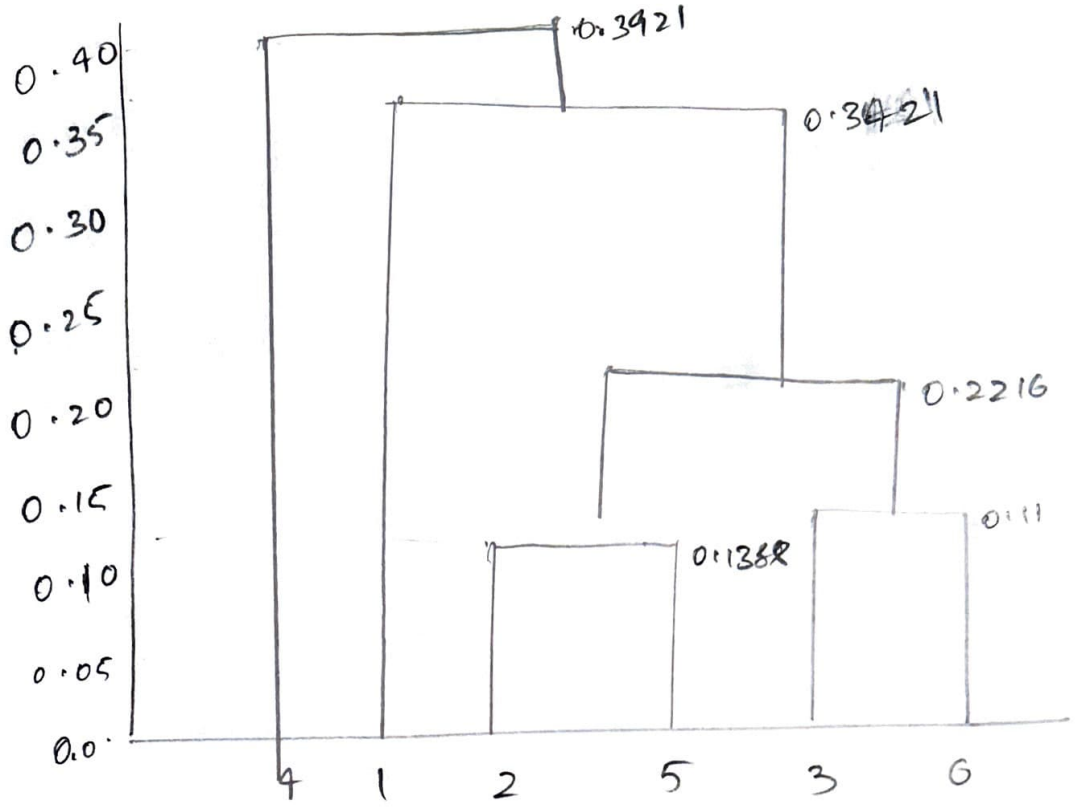
$\Rightarrow 0.3921$

Final table is

	P_2 P_5 P_1	P_3 P_6 P_4
P_2 P_5 P_1	0	
P_3 P_6 P_4	0.3921	0

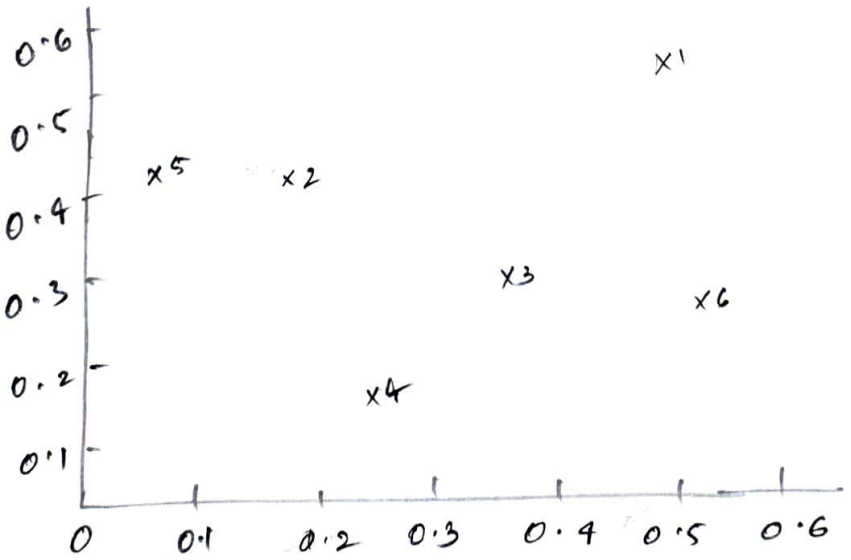
Final cluster is





(3) Average link proximity.

Graphical representation of the given points.



The minimum value in the lower bound is 0.1100 i.e, b/w P_3 & P_6

The distance matrix is

$$Avg[dist(P_3 P_6), P_1)]$$

$$dist((P_3 P_6) P_1)] \Rightarrow \frac{1}{2}[dis(P_3 P_1) + dist(P_6 P_1)]$$

$$\Rightarrow \frac{1}{2}[0.2218 + 0.2347]$$

$$\Rightarrow 0.2282$$

Distance with $P_2 \Rightarrow Avg[dist(P_3 P_6), P_2)]$

$$\Rightarrow \frac{1}{2}[dist(P_3 P_2) + dist(P_6 P_2)]$$

$$\Rightarrow \frac{1}{2}[0.1483 + 0.2540]$$

$$\Rightarrow 0.2011$$

Distance with $P_4 \Rightarrow Avg[dist(P_3 P_6), P_4)]$

$$\Rightarrow \frac{1}{2}[dist(P_3 P_4) + dist(P_6 P_4)]$$

$$\Rightarrow \frac{1}{2}[0.1523 + 0.2216)]$$

$$\Rightarrow 0.1864$$

Distance with $P_5 \Rightarrow Avg[dist(P_3 P_6), P_5]$

$$\Rightarrow \frac{1}{2}[dist(P_3 P_5) + dist(P_3, P_5)]$$

$$\Rightarrow \frac{1}{2}(0.2843 + 0.3921)$$

$$\Rightarrow 0.3382.$$

One updated distance matrix for cluster $(P_2 P_6)$ is

	P_1	P_2	$P_3 P_6$	P_4	P_5
P_1	0				
P_2	0.2357	0			
$P_3 P_6$	0.2282	0.2011	0		
P_4	0.3688	0.2042	0.1864	0	
P_5	0.3421	0.1388 (Min)	0.3382	0.2932	0

Now, the minimum value is 0.1388 b/w P_2 & P_5, the distance

Distance of $(P_2 P_5)$ with P_1 $\Rightarrow$ Avg[dist$((P_2 P_5) P_1)$]

$\Rightarrow$ ½ [dist(P_2, P_1) + dist(P_5, P_1)]

$\Rightarrow$ ½ [0.2357 + 0.3421]

$\Rightarrow$ 0.2889

Distance with $(P_3 P_6)$ $\Rightarrow$ ½ [dist[$(P_2 P_5)(P_3 P_6)$]]

$\Rightarrow$ ½ [0.2011 + 0.3382]

$\Rightarrow$ 0.2696

Dist with P_4 $\Rightarrow$ Avg(dist$(P_2 P_5) P_4$)

$\Rightarrow$ ½ [dist$(P_2 P_4)$ + dist(P_5, P_4)]

$\Rightarrow$ ½ [0.2042 + 0.2932] $\Rightarrow$ 0.2487

Updated distance table is for (P_2P_5)

	P_1	P_2P_5	P_3P_6	P_4
P_1	0			
P_2P_5	0.2889	0		
P_3P_6	0.2282	0.2696	0	
P_4	1.3688	0.2487	0.1864	0

Next minimum value in 0.1864 b/w P_4 & P_3, P_6

Distance with P_1 for (P_3P_6) P_4 in

$\Rightarrow$ Avg(dist$(P_3, P_6, P_4), P_1)$

$\Rightarrow \frac{1}{2}[\text{dist}(P_3P_6)P_1 + \text{dis}(P_4, P_1)]$

$\Rightarrow \frac{1}{2}[0.2282 + 0.3688]$

$\Rightarrow 0.2985$

Distance to $(P_2P_5) \Rightarrow \frac{1}{2}[\text{dist}(P_3P_6)(P_2P_5) + \text{dist}(P_4, (P_2P_5))]$

$\Rightarrow \frac{1}{2}[0.2696 + 0.2487]$

$\Rightarrow 0.2591$

Updated distance table for P_4 & (P_3, P_6) is

	P_1	$P_2 P_5$	$P_3 P_6 P_4$
P_1	0		
$P_2 P_5$	0.2879	0	
$P_3 P_6 P_4$	0.2985	0.2591	0

Minimum

The minimum value i.e) 0.2591 b/w $P_2 P_5$ & $P_3 P_6 P_4$

Distance from P_1 to $(P_2 P_5)$ $(P_3 P_6, P_4)$

2) $Avg[dist((P_2 P_5)(P_3 P_6 P_4)) P_1))]$

$\Rightarrow \frac{1}{2}[dist(P_3, P_6, P_4) P_1)) + ((P_2 P_5), P_1)]$

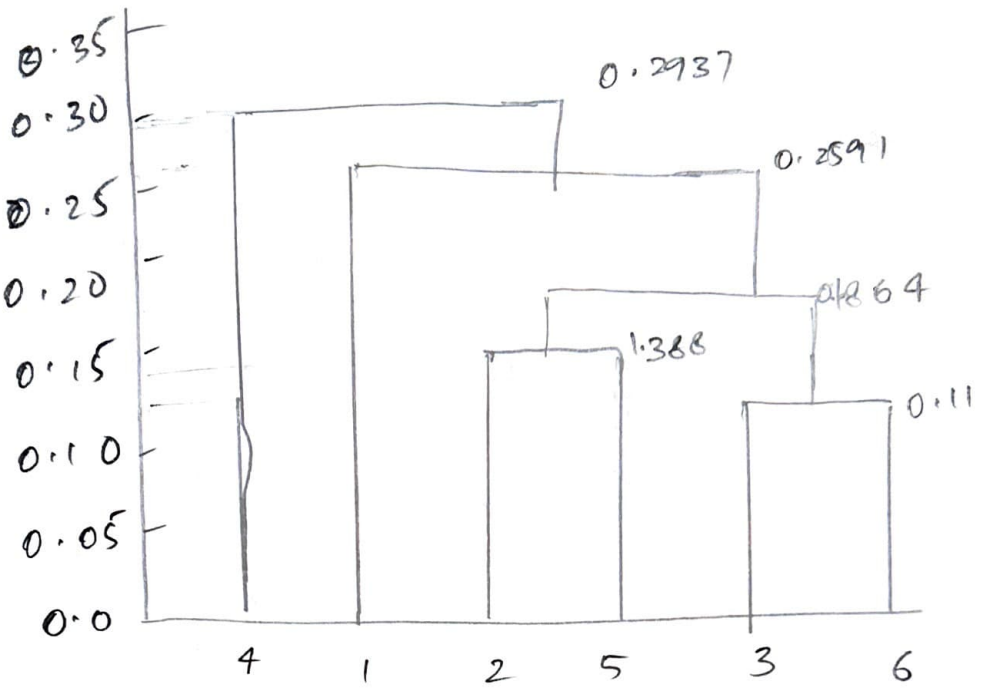
$\Rightarrow \frac{1}{2}[0.2985 + 0.2889]$

$\Rightarrow 0.2937$

The distance table after merging is

	P_1	$P_2 P_5 P_3 P_4$
P_1	0	
$P_2 P_5 P_3 P_6 P_4$	0.2937	0

The final cluster dendrogram is :



Cluster after applying average link.