Assignment-6

Sowjanya Sunkavalli
700731896

① Calculate the clustering representations and dendogram using Single, Complete and Average link proximity function in hierarchical clustering technique.

(a). Single link proximity function.

Given below in the table with euclidian distance between each individual points.

	P_1	P_2	P_3	P_4	P_5	P_6
P_1	0	0.2357	0.2218			
P_2	0.2357	0	0.1483			
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

Table ①

Considering lower bound values, since upper bound value are equal to lowerbound and finding the cluster according to that.

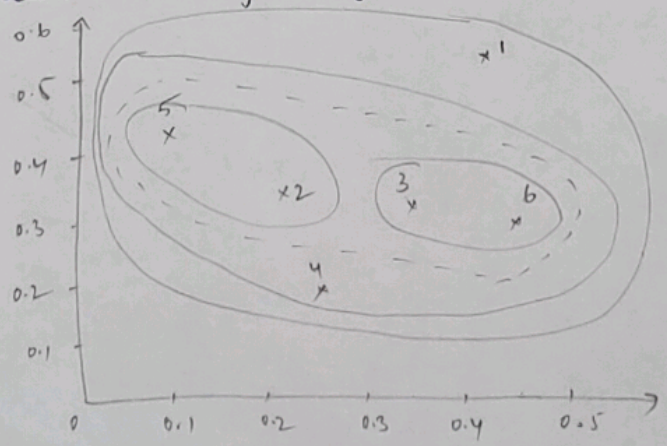


fig ①

Graphical representation of the given six points.

after finding the euclidean distance between each individual points, next step is

⇒ Merging the two closest members and updating the distance table.

① From the above table min. value is 0.1100 i.e between P_3 and P_6. after merging two member we need to find the distance with other member usig the formula. $\min[dist(clusters, otherpoints)]$

⇒ Distance with P_1 ⇒ $\min[dist((P_3, P_6), P_1)]$

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

$= \min(0.2218, 0.2347) \Rightarrow \underline{0.2218}$

⇒ Distance with P_2 ⇒ $\min[dist((P_3, P_6), P_2)]$

$= \min(dist(P_3, P_2), (P_6, P_2))$

$= \min(0.1483, 0.2540) \Rightarrow \underline{0.1483}$

⇒ Distance with P_4 ⇒ $\min[dist((P_3, P_6), P_4)]$

$= \min((P_3, P_4), (P_6, P_4))$

$= \min(0.1513, 0.2216) \Rightarrow \underline{0.1513}$

⇒ Distance with P_5 ⇒ $\min[dist((P_3, P_6), P_5)]$

$= \min((P_3, P_5), (P_6, P_5))$

$= \min(0.2843, 0.3921) \Rightarrow \underline{0.2843}$.

So updated distance table after merging P_3 and 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	0.2218	0.1483	0		
P_4	0.3688	0.2042	0.1513	0	
P_5	0.3421	0.1388 (min)	0.2843	0.2932	0

(P_3, P_6]

(P_2, P_5]

② From the above table which is updated 0.1388 & 0.1483 is the min. value. then join P_2P_6 and P_2 and P_5.

Distance of (P_2 P_5) with P_1

⇒ Min (dist ((P_2P_5) P_1))

= min ((P_2 P_1) (P_5 P_1))

= min (0.2357, 0.3421) = 0.2357 ⇒ P_1

Distance of (P_2 P_5) with P_3P_6

= min (dist ((P_2 P_5) (P_3 P_6)) ⇒ min ((P_2(P_3 P_6)) (P_5(P_3 P_6)))

= (0.1483, 0.2843) ⇒ 0.1483 ⇒ (P_3 P_6)

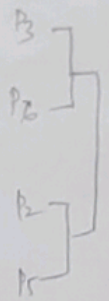
Distance of (P_2 P_5) with P_4

= min (dist ((P_2 P_5) P_4)) ⇒ min (dist ((P_2 P_4) (P_5 P_4)))

= (0.2042, 0.2932) ⇒ 0.2042 ⇒ P_4

updated distance table after merging P_2 with P_5 in

	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 (min)	0	
P_4	0.3688	0.2042	0.1513	0



since the min. value is 0.1483, so merge (P_2P_5) and (P_3P_6) and update distance matrix.

⇒ Distance of $(P_2P_5)(P_3P_6)$ with P_1

⇒ $\min(dist((P_2P_5)(P_3P_6), P_1))$

$= \min(dist((P_2P_5)P_1)((P_3P_6)P_1))$

$= \min(0.2357, 0.2218) \Rightarrow \underline{0.2218} \Rightarrow P_1$

⇒ Distance of $(P_2P_5)(P_3P_6)$ with P_4

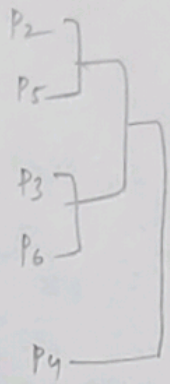
⇒ $\min(dist((P_2P_5)(P_3P_6)P_4))$

$= \min(dist((P_2P_5)P_4)((P_3P_6)P_4))$

$= \min(0.2042, 0.1513) \Rightarrow \underline{0.1513} \Rightarrow P_4$

So, the updated distance table is as shown below.

	P_1	$P_2P_5P_3P_6$	P_4
P_1	0		
P_2P_5, P_3P_6	0.2218	0	
P_4	0.3688	0.1513 (min)	0



Merge P_4 and P_2P_5, P_3P_6 and update the distance table.

Distance with P_1 ⇒ $\min[dist((P_2P_5P_3P_6), P_4P_1))]$

$= \min[dist((P_2P_5P_3P_6)P_1)(P_4P_1))]$

$= (0.2218, 0.3688) \Rightarrow 0.2218$

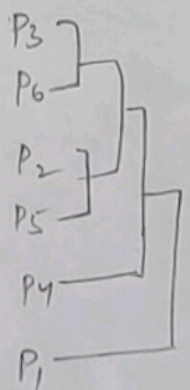
Distance with

	P_1	$P_2P_5P_3P_6$	P_4
P_1	0		
$P_2P_5P_3P_6P_4$	0.2218 (min)	0	

⇒ Now $\min(dist((P_2P_5P_3P_6P_4)(P_1)))$

⇒ 0.22.

So, final cluster is shown in figure ① and below is the dendogram



fig②

② Complete link prominity function

for Complete link prominity we need merge the closest points and update the distance table with below formula according to the below formula.

Distance ⇒ Max (dist (clustering point, points))

Graphical representation of given points

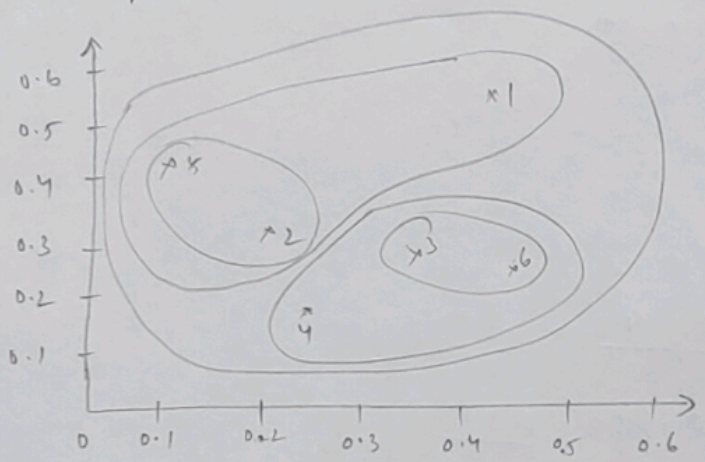


fig ①

From the table ① min. value is 0.1100 ; ie merge P_3 and P_6 and update the distance table.

Distance with P_1 ⇒ Max (dist (P_3 P_6) P_1))

⇒ Max (dist (P_3 P_1) (P_6 P_1))

= max (0.2218, 0.2347) ⇒ 0.2347

P_3
P_6

Distance with P_2 ⇒ Max (dist ((P_3 P_6) P_2))

- Max (dist ((P_3 P_2) (P_6 P_2)))

= max (0.1483, 0.2540) ⇒ 0.2540

Distance with P_4 ⇒ Max(dist((P3P6) P4))

= Max(0.1513, 0.2216) = 0.2216

Distance with P_5 ⇒ Max(dist((P3 P6) P5))

= Max(0.2843, 0.3921) ⇒ 0.3921

So, the updated distance table after merging P_3 and P_6 is

	P1	P2	P3 P6	P4	P5
P1	0				
P2	0.2357	0			
P3 P6	0.2347	0.2540	0		
P4	0.3688	0.2042	0.2216	0	
P5	0.3421	0.1388 (Min.)	0.3921	0.2932	0

P_3, P_6]

P_2, P_5]

Min. value is 0.1388, merge P_2, P_5 and update the distance table.

distance of $(P_2 P_5)$ with P_1 ⇒ Max(dist((P2 P5) P1))

= Max(0.2357, 0.3421) = 0.3421 P_1

Distance of $(P_2 P_5)$ with $P_3 P_6$ ⇒ Max(dist((P2 P5)(P3 P6)))

= Max(0.2540, 0.3921)

= 0.3921 → $P_3 P_6$

Distance of (P_2P_5) with P_4

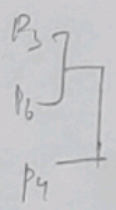
$\Rightarrow$ Max (dist $((P_2P_5)P_4)$) $\rightarrow$ Max $(0.2042, 0.2932)$

$= 0.2932 \Rightarrow P_4$

Updated distance table after merge P_2 and P_5 in

P_2P_6 $\rightarrow$ P_4

	P_1	P_2P_5	
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 (min) 0



Distance $\Rightarrow$ of $P_4(P_2P_6)$ with P_1

$\Rightarrow$ Max (dist $((P_2P_6)P_4)P_1)$)

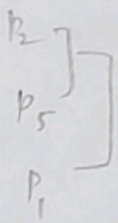
$=$ Max $(0.2347, 0.3688) \Rightarrow P_1 \Rightarrow 0.3688$

Distance with P_2P_5 $\Rightarrow$ Max $((P_4(P_3P_6))(P_2P_5))$

$=$ Max $(0.2932, 0.3921) \Rightarrow P_2P_5 \Rightarrow 0.3921$

	P_1	P_2P_5	$P_3P_4P_6$
P_1	0		
P_2P_5	0.3421 (min)	0	
$P_3P_6P_4$	0.3688	0.3921	0

Merge P_2P_5 and P_1



Distance with $P_3P_6P_4$

$= \text{Dist}(\text{Max}((P_1(P_2P_5))((P_3P_6)P_4)))$

$= \text{Max}(0.3688, 0.392))$

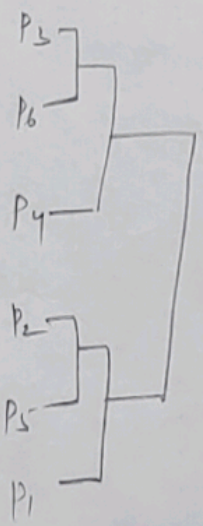
$= 0.392).$

Final table is

	$P_2P_5P_1$	$P_3P_6P_4$
$P_2P_5P_1$	0	
$P_3P_6P_4$	0.3921	0

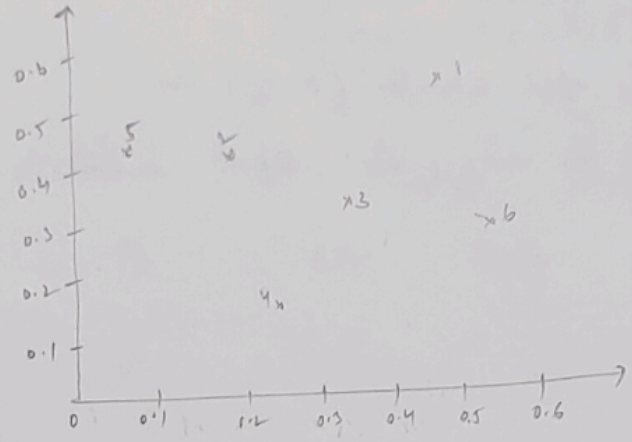
Final clustering is shown in figure ① of ②

and dendogram is



(2) Average link

Graphical representation of the points given.



the minimum value in the lower bound is 0.1100 i.e between P_3 and P_6.

$\begin{bmatrix} P_3 \\ P_6 \end{bmatrix}$

The distance matrix is $Avg\,[dist((P_3P_6), P_1)]$

$$dist((P_3P_6), P_1) = \frac{1}{2}\left[dist(P_3P_1) + dist(P_6, P_1)\right]$$

$$= \frac{1}{2}\left[0.2218 + 0.2347\right]$$

$$= 0.2282$$

Distance with P_2 → $Avg\,[dist(P_3P_6), P_2)]$

$$= \frac{1}{2}\left[dist(P_3P_2) + dist(P_6, P_2)\right]$$

$$= \frac{1}{2}\left[0.1483 + 0.2540\right]$$

$$= 0.2011$$

Distance with P_4 → $Avg\,[dist(P_3P_6), P_4)]$

$$= \frac{1}{2}\left[dist(P_3P_4) + dist(P_6P_4)\right] \Rightarrow \frac{1}{2}\left[(0.1512 + 0.2216)\right]$$

$$= 0.1864$$

Distance with $P_5 \Rightarrow Avg[dist(P_3P_6), P_5]$

$= Avg[dist(P_3, P_5) + dist(P_3, P_5)]$

$= 1/2(0.2843 + 0.3921) \Rightarrow 0.3382.$

Updated distance matrix for cluster $(P_3 P_6)$ is

	P_1	P_2	$P_3 P_6$	P_4	P_5
P_1	0.00				
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	0.2922	0

Now, the minimum value is 0.1388 between P_2 and P_5

Update the distance,

P_3]
P_6]

Distance of $(P_2 P_5)$ with $P_1 \Rightarrow$

P_2]
P_5]

$= Avg[dist(P_2P_5)P_1)]$

$= 1/2[dist(P_2, P_1) + dist(P_5, P_1)]$

$= 1/2[0.2357 + 0.3421]$

$= 0.2889$

Distance with $(P_3 P_6) \Rightarrow 1/2[dist[(P_2P_5)(P_3P_6)]]$

$= 1/2[0.2011 + 0.3382] \Rightarrow 0.2696$

Distance with P_4 $\Rightarrow$ $Avg(dist(P_2P_5), P_4)$

$= 1/2[dist(P_2P_4) + dist(P_5, P_4)]$

$= 1/2[0.2042 + 0.2932] \Rightarrow 0.2487$

Updated distance table in for (P_2P_5)

	P_1	P_2P_5	P_3, P_6	P_4
P_1	0			
P_2P_5	0.2889	0		
P_3P_6	0.2282	0.2696	0	
P_4	0.3688	0.2487	0.1864 (min.)	0

Next minimum value in 0.1864 between P_4 and P_3, P_6

Distance with P_1 for $(P_3, P_6)P_4$ in

$\Rightarrow Avg(dist(P_3P_6P_4), P_1))$

$= 1/2[dist(P_3P_6)P_1 + dist(P_4, P_1)]$

$= 1/2[0.2282 + 0.3688]$

$= 0.2985$

P_3, P_6, P_4

P_5, P_5

Distance to (P_2P_5) $\Rightarrow$ $1/2[dist(P_3P_6)(P_5P_5) + dist(P_4, (P_2, P_5))]$

$= 1/2[0.2696 + 0.2487] \Rightarrow 0.2591$

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

	P_1	P_2, P_5	P_3, P_6, P_4
P_1	0		
$P_2 P_5$	0.2819	0	
P_3, P_6, P_4	0.2985	0.2591 (min.)	0

Now, the minimum value is 0.2591 between P_2, P_5 and P_3, P_6, P_4

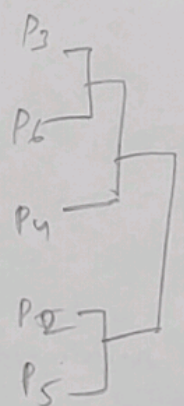
Distance from P_1 to $(P_2 P_5)(P_3, P_6, P_4)$

$$= \text{Avg}\left[dist\left((P_2 P_5)(P_3, P_6, P_4)\right) P_1)\right]$$

$$= 1/2\left[dist(P_3, P_6, P_4), P_1)) + ((P_2, P_5) P_1)\right]$$

$$= 1/2\left[0.2985 + 0.2989\right]$$

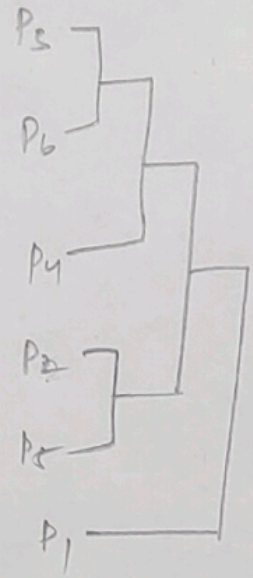
$$= 0.2937.$$



Updated Updated distance table after merging is

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

then final cluster or dendogram is as follows



Cluster representation of the given points after applying Average link function is as follows.