Assignment - 6 K. Lavanya (700728999)

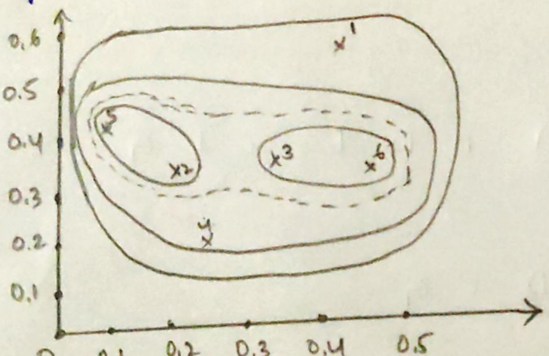
1) Calculate the clustering representation and dendogram using single, Complete and Average link priority function in hierarchical clustering technique.

a) Single link priority function.

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

	P1	P2	P3	P4	P5	P6
P1	0					
P2	0.2357	0				
P3	0.2218	0.1483	0			
P4	0.3688	0.2042	0.1513	0		
P5	0.3421	0.1388	0.2842	0.2932	0	
P6	0.2347	0.2540	0.1100	0.2216	0.3921	0

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



Graphical representation of the given six points.

after find the euclidean distance b/w 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 in between P_3 and P_6 after merging two member we need to find the distance with other member using the formula min[dist (Cluster, other points)]

→ 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)

= 0.2218.

①

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)

= 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)

= 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)

= 0.2843

So updated distance table after merging P_3 and P_6 is

	P_1	P_2	$P_3\ P_6$	P_4	P_5
P_1	0				
P_2	0.2357	0			
$P_3\ P_6$	0.2218	0.1483	0		
P_4	0.3688	0.2042	0.1513	0	
P_5	0.3421	0.1388 min	0.2843	0.2932	0.

(2) from the above table which is updated 0.1388 is the min value then join P_2 and P_5

Distance of $(P_2 P_5)$ with P_1 = min (dist $((P_2 P_5) P_1)$) $\left.\begin{matrix} P_2 \\ P_5 \end{matrix}\right]$

= 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 $P_3 P_6$ = 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)

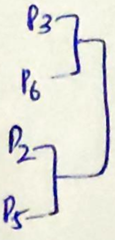
= 0.1483 ⇒ $(P_3 P_6)$

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 mergin 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 min	0	
P_4	0.3688	0.2042	0.1513	0



Since the min value is 0.1483 merge (P_2P_5) and (P_3P_6) and update distance matrix

Distance of $(P_2P_5)(P_3P_6)$ with $P_1 = \min(dist((P_2P_5)(P_3P_6)P_1)))$

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

$= \min(0.2357, 0.2218)$

$= 0.2218 \rightarrow P_1$

Distance of $(P_2P_5)(P_3P_6)$ with $P_4 = \min(dist((P_2P_5)(P_3P_6)P_4)))$

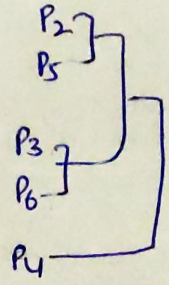
$= \min(dist((P_2P_5)P_4)((P_3P_6)P_4)))$

$= \min(0.2042, 0.1513$

$= 0.1513 \Rightarrow P_4$

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_2, P_5, P_3, P_6 and update the distance table.

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

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

$= (0.2218, 0.3688) = 0.2218$

	P_1	$P_2P_5P_3P_6P_4$
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

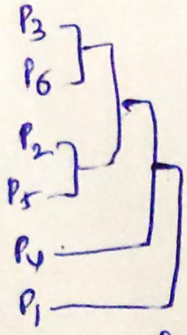


fig ②

② complete link proximity function.

For complete link proximity we need merge the closest point and update the distance table with below ~~function~~ according to below formula.

Distance = Max (dist (cluster point, points))

Graphical representation of given point

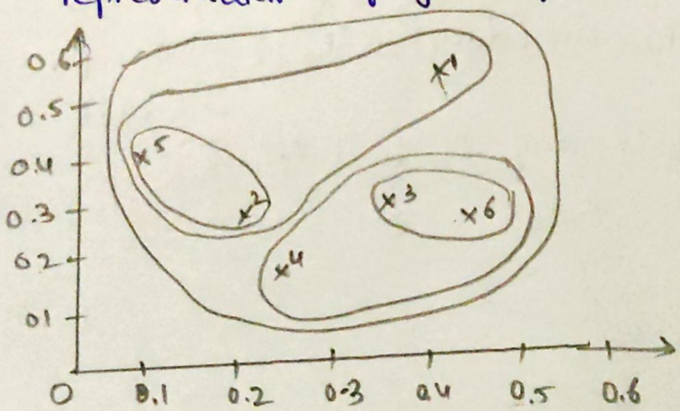


fig ①

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

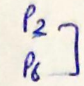
Distance with P_2 ⇒ max (dist $(P_3 P_6)\ P_2)$) = max (dist $((P_3 P_2)(P_6 P_2))$)

= max (0.1483, 0.2540) = 0.2540

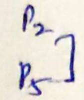
Distance with P_4 = max (dist $((P_3 P_6)\ P_4)$) = max (0.1513, 0.2216) = 0.2216

Distance with P_5 = max (dist $(P_3 P_6) P_5)$) = Max(0.2843, 0.3921) = 0.3921

So, the 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	0.2347	0.2540	0		
P_4	0.3688	0.2042	0.2216	0	
P_5	0.3421	0.1388 min	0.3921	0.2932	0



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

Distance of (P_2P_5) with P_1 = max (dist $((P_2P_5)\ P_1)$) = max $(0.2357, 0.3421)$
$= 0.3421 \rightarrow P_1$

Distance of (P_2P_5) with P_3P_6 ⇒ Max (dist $(P_2P_5)(P_3P_6)$)) = max $(0.2540, 0.3921)$
$= 0.3921 \rightarrow P_3P_6$

Distance of (P_2P_5) with P_4 = Max (dist $(P_2P_5)P_4$)) = max $(0.2042, 0.2932)$
$= 0.2932 \rightarrow P_4$.

updated distance table after merg P_2 and P_5 is

	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 min	0

Distance of $P_4(P_3P_6)$ with P_1 = max (dist $((P_3P_6)P_4)P_1$)) = max $(0.2347, 0.3688)$
$= 0.3688 \Rightarrow P_1$

Distance with P_2P_5 = max $((P_4(P_3P_6))(P_2P_5))$) = max $(0.2932, 0.3921) = 0.3921$
$\Rightarrow P_2P_5$

	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

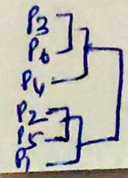
merge P_2P_5 and P_1

Distance with $P_3P_6P_4$ = Dist (max $((P_1(P_2P_5))((P_3P_6)P_4))$) = max$(0.3688, 0.3921)$
$= 0.3921$

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 fig ① of ①
and dendogram is



5

The minimum value in lower bound 0.1100 b/w P_3 & P_6
first cluster $\boxed{3 \; 0 \; 6}$

The distance matrix is $AVG\,[dist\,(P_3, P_6), P1] = dis\,((P_3, P_6, P1))$
$= \frac{1}{2}\,(dis(P_3, P_1) + dis\,(P6, P1))$
$= \frac{1}{2}\,(0.2218 + 0.2347)$
$= 0.2282$

The distance matrix is, $AVG\,[dist\,(P_3, P_6), P_2] = dis\,((P_3 P_6)\,P_2)$
$= \frac{1}{2}\,[dis\,(P_3, P_2) + dis\,(P_6, P_2))$
$= \frac{1}{2}\,((0.1482) + (0.2540))$
$= 0.2011$

The distance matrix is $AVG[(dis\,(P_3, P_6), P_4)] = dis\,((P_3, P_6); P_4)$
$= \frac{1}{2}\,(dist\,(P_3, P_4) + dis\,(P_6, P_4)]$
$= \frac{1}{2}\,[0.1513 + 0.2216]$
$= 0.1864$

The distance matrix is, $AVG[dist(P_3, P_6), P_5] = dis\,((P_3, P_6)\,P_5)$
$= \frac{1}{2}\,(dis\,(P_3, P_5) + dis\,(P_6, P_5))$
$= \frac{1}{2}\,(0.2843 + 0.3921)$
$= 0.3382$

The updated distance matrix for cluster (P_3, P_6)

	P1	P2	P3 P6	P4	P5
P1	0.0000				
P2	0.2357	0.			
P3, P6	0.2282	0.2011	0		
P4	0.3688	0.2042	0.1864	0	
P5	0.3421	0.1388	0.3382	0.2932	0

The min value is 0.1388 b/w P_2 & P_5
second cluster $\boxed{3 \; 1 \; 6}$ $\boxed{2 \; 2 \; 5}$

To update the distance matrix, $AVG\,(dist\,(P_2, P_5), P1)$
$dis((P_2, P_5), P1) = \frac{1}{2}\,(dis(P_2, P_1) + dis\,(P_5, P_1)) = \frac{1}{2}\,(0.2357 + 0.3421)$
$= 0.2889$

To update the distance matrix, $AVG\,(dist\,((P_2, P_5), (P_3, P_6)))$
$dis\,((P_2, P_5), (P_3, P_6)) = \frac{1}{2}\,(dis\,(P_2, (P_3, P_6)) + dis\,(P_5, (P_3, P_6))$
$= \frac{1}{2}\,(0.2011 + 0.3382) = 0.2696$

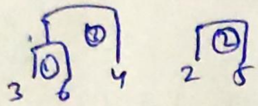
To update the distance matrix, $AVG\,(dist\,(P_2, P_5), P_4)$
$dist\,((P_2, P_5), (P_4)) = \frac{1}{2}\,(dist\,(P_2, P_4) + dis\,(P_5, P_4)) = \frac{1}{2}\,(0.2042 + 0.2932)$
$= 0.2487$

The updated distance matrix for cluster (P_2, P_5)

	P_1	P_2, P_5	P_3, P_6	P_4
P_1	0			
P_2, P_5	0.2889	0		
P_3, P_6	0.2282	0.2696	0	
P_4	0.3688	0.2487	0.1864	0

The min value in lower bound is 0.1864 b/w P_4 & P_3, P_6

Third cluster



To update the distance matrix, AVG (dist (P_3, P_6, P_4), P_1) = dist((P_3, P_6, P_4), P_1)

$= \frac{1}{2}$ (dist ((P_3, P_6) P_1) + (P_4, P_1))]

$= \frac{1}{2}$ [(0.2282) + (0.3688)]

$= 0.2985$

To update the distance matrix, AVG (dist (P_3, P_6, P_4), (P_2, P_5))

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

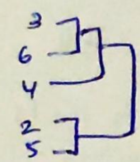
$= \frac{1}{2}$ (0.2696 + 0.2487) = 0.2591

updated matrix for cluster (P_4 ; (P_3 & P_6))

	P_1	P_2, P_5	P_3, P_6, P_4
P_1	0		
P_2, P_5	0.2889	0	
P_3, P_6, P_4	0.2985	0.2591	0

The min value is 0.2593 b/w P_2, P_5 & P_3, P_6, P_4

fourth cluster.



To update the distance matrix, AVG (dist (P_3, P_6, P_4), (P_2, P_5)) P_1)

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

$= \frac{1}{2}$ (0.2985 + 0.2889)

$= 0.2937.$

	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

The last cluster P_2, P_5, P_3, P_6, P_4 with P_1