Assignment-6

Name: JONNALAGADDA CHAKRADHAR.

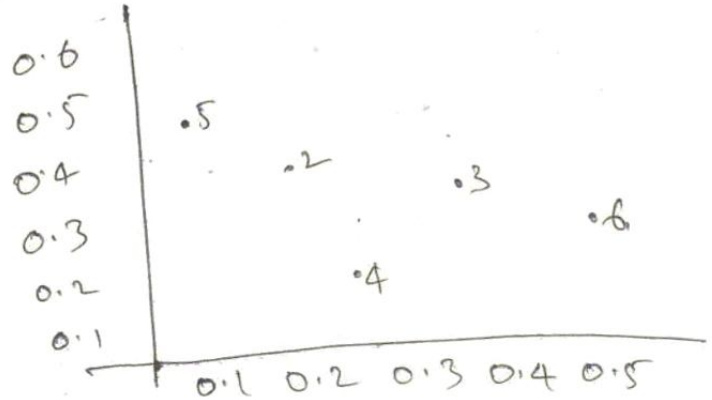
Student ID: 700744495
Network ID: cxj4950.

Q1) Find out clustering representations, Dendrogram using single, complete and Average link proximity function in Hierarchial clustering technique.

point	x-coordinate	y-coordinate.
P_1	0.4005	0.5306
P_2	0.2148	0.3854
P_3	0.3457	0.3156
P_4	0.2652	0.1875
P_5	0.0789	0.4139
P_6	0.4548	0.3022

	P_1	P_2	P_3	P_4	P_5	P_6
P_1	0.0000	0.2357	0.2218	0.3688	0.3421	0.2347
P_2	0.2357	0.0000	0.1483	0.2042	0.1388	0.2540
P_3	0.2218	0.1483	0.0000	0.1513	0.2843	0.1100
P_4	0.3688	0.2042	0.1513	0.0000	0.2932	0.2216
P_5	0.3421	0.1388	0.2843	0.2932	0.0000	0.3921
P_6	0.2347	0.2540	0.1100	0.2216	0.3921	0.0000

By single link:- In this, the proximity of two clusters is minimum of the distance b/w any two points in 2 diff clusters.
The single link technique is good for non elliptical shapes, but sensitive to noise & outliers



→From table I we can observe the distance between P_3 & P_6 is 0.11

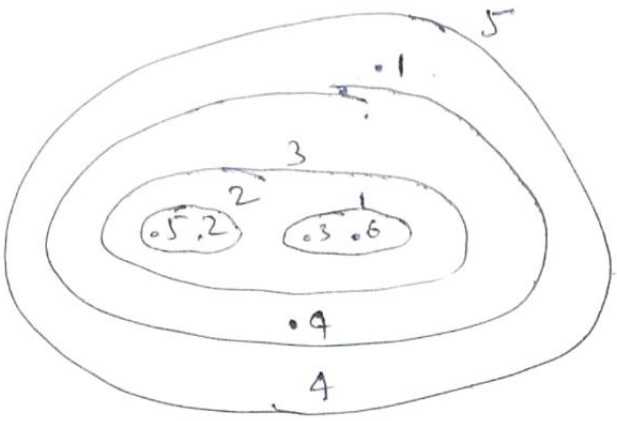
→The height at which two clusters are merged can be represented as distance between two clusters.

Distance between clusters {3,6} & {2,5} is given by.

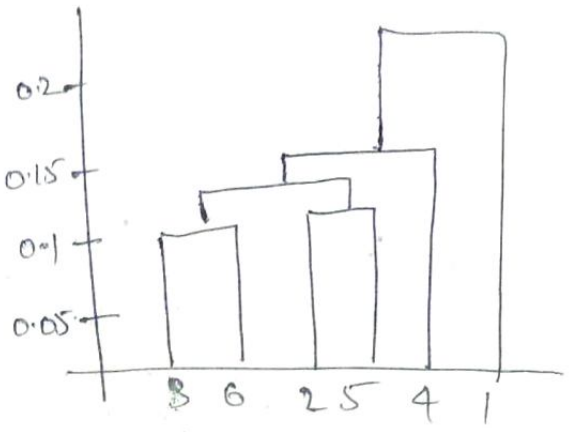
dist({3,6},{2,5}) = min(dist(3,2), dist(6,2), dist(3,5), dist(6,5))

⇒ min (0.15, 0.25, 0.28, 0.39)

⇒ 0.15



Single link clustering



Single link dendogram

Complete link: The proximity of two clusters is defined as "the maximum of the distance between any two points in two different clusters.

→This is less susceptible to noise & outliers, but it can break large clusters and its favours globular shapes.

→Here points 3 and 6 are merged first.

{3,6} is merged with {4} instead of {2,5} or {1} this is because dist({3,6},{4}) = max(dist(3,4), dist(6,4))

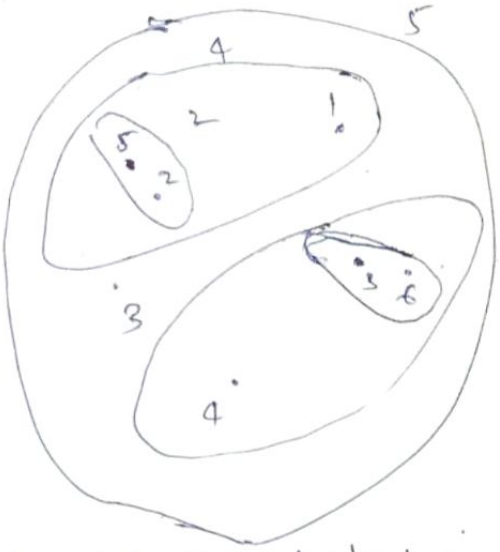
= max(0.15, 0.22)

= 0.22.

dist({3,6},{2,5}) = max(0.15, 0.25, 0.28, 0.39)

= 0.39.

dist({3,6},{1}) = max(dist(3,1), dist(6,1))
= max(0.22, 0.23)
= 0.23



Complete link clustering

complete link dendrogram

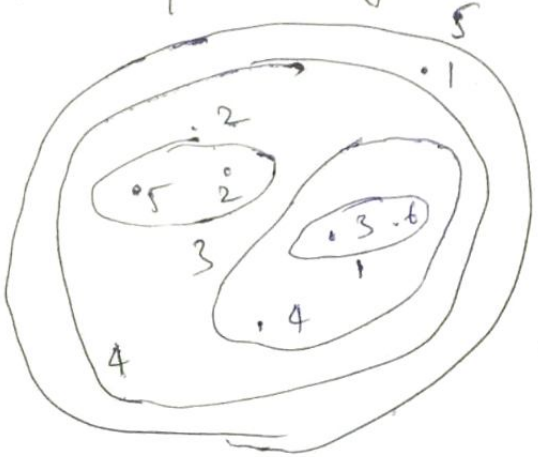
Average link: The average group approach is applied to the points.

dist({3,6,4},{1}) = (0.22+0.37+0.23) / (3×1)
= 0.28.

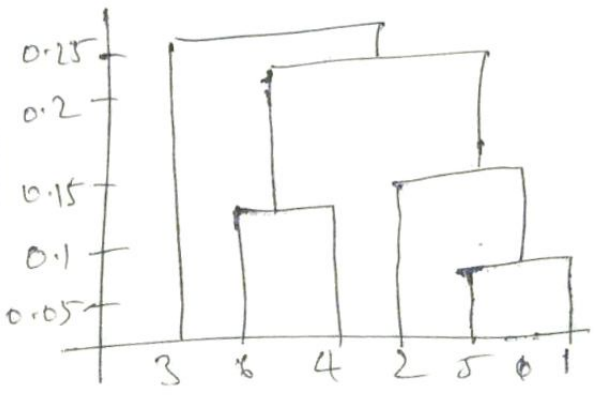
dist({2,5},{1}) = (0.24+0.34) / (2×1)
= 0.29.

dist({3,6,4},{2,5}) = (0.15+0.28+0.25+0.39+0.20+0.29) / (3×2)
= 0.26.

Here, because dist({3,6,4},{2,5}) is smaller than dist({3,6,4},{1}) and dist({2,5},{1}) clusters {3,6,4} and {2,5} are merged at the fourth stage.



Average link clustering.



Average link dendogram