No: 700739768

Name: Uday Kumar Kuppam

Assignment-6

USER ID: UXK97680

11/15/2022

Question -1

Find out clustering representations, & Dendrogram using Single, Complete and Average link proximity function in Hierarchial clustering technique ?

point	x-coordinate	y co-ordinate
P1	0.04005	0.5306
P2	0.2148	0.3854
P3	0.3457	0.3156
P4	0.2652	0.1875
P5	0.0789	0.4139
P6	0.4548	0.3022

~~Fig.~~ Table 1

X-Y coordinates

Distance Matrix

	P1	P2	P3	P4	P5	P6
P1	0.000	0.2357	0.2218	0.3688	0.3421	0.2347
P2	0.2357	0.000	0.1483	0.2042	0.1388	0.2540
P3	0.2218	0.1483	0.000	0.1513	0.2843	0.1100
P4	0.3688	0.2042	0.1513	0.0000	0.2932	0.2216
P5	0.3421	0.1388	0.2843	0.2932	0.0000	0.3921
P6	0.2347	0.2540	0.1100	0.2216	0.3921	0.0000

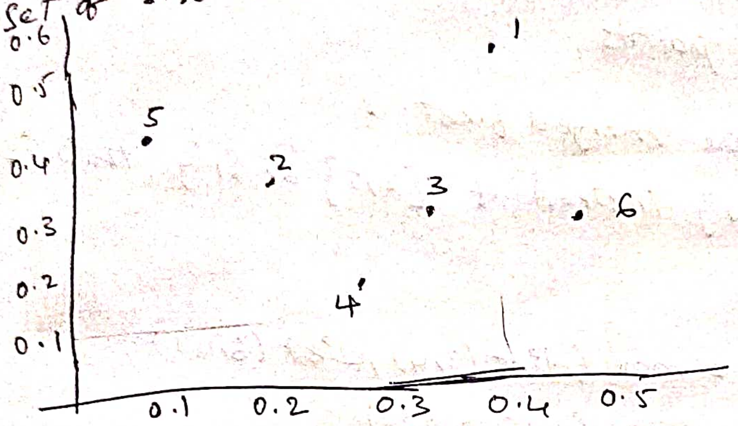
~~Fig.~~ Table: 2

By-single link :

* For single link hierarchial clustering, the proximity of two clusters is minimum of the distance between any two points in 2 different clusters.

* the single link technique is good for non elliptical shapes, but sensitive ~~for~~ to noise & outliers

* Applying single link technique to our Example data set of six points set of six - 2 dimensional points



→ from table 1, we can obseve distancance between P3 & P6 is 0.11.

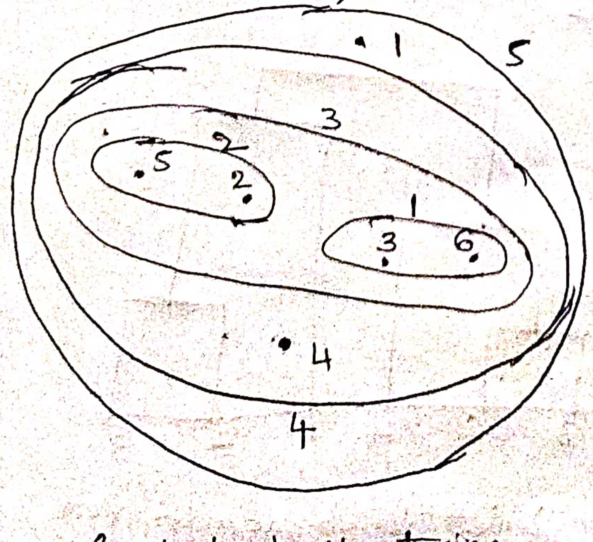
→ the height at which two clusters are ~~are~~ merged ~~is~~, ~~the~~ can be represented as ~~is~~ distance between two clusters.

distance between clusters {3,6} & {2,5} is Given by

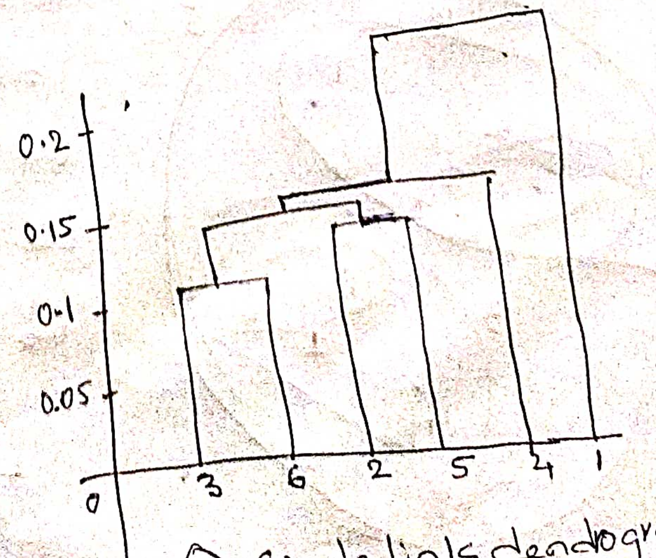
$dist(\{3,6\},\{2,5\}) = min(dist(3,2), dist(6,2), dist(3,5)\ \&\ dist(6,5))$

$\Rightarrow min(0.15, 0.25, 0.28, 0.39)$

$\Rightarrow 0.15.$



Single link Clustering



(b) Single link dendrogram

Complete link

→ In Complete link of hierorchial clustering, the Proximity of two Clusters is defined as "the maximum of the distance between any two points in two different clusters.

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

→ Below fig shows results of Applying Max to the sample data set of six points

→ Here points 3 and 6 are merged first. {3, 6} is merged With {4} instead of {2,5} or {1} this is

Because

$$\text{dist}(\{3,6\}, \{4\}) = \max(\text{dist}(3,4), \text{dist}(6,4))$$

$$= \max(0.15, 0.22)$$

$$= 0.22$$

$$\text{dist}(\{3,6\},\{2,5\}) = \max(\text{dist}(3,2), \text{dist}(6,2), \text{dist}(3,5), \text{dist}(6,5))$$

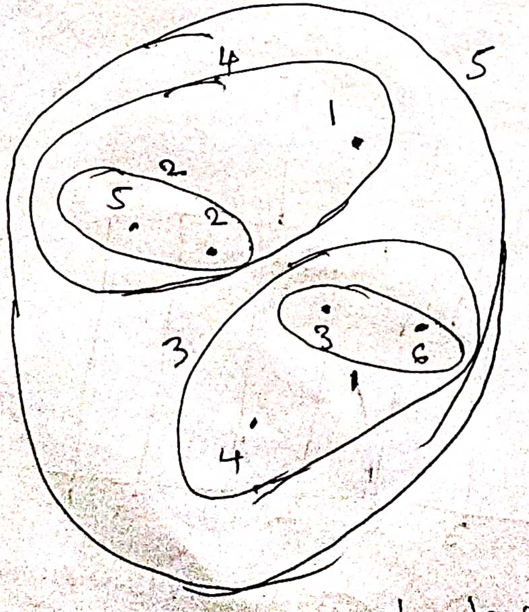
$$= \max(0.15, 0.25, 0.28, 0.39)$$

$$= 0.39$$

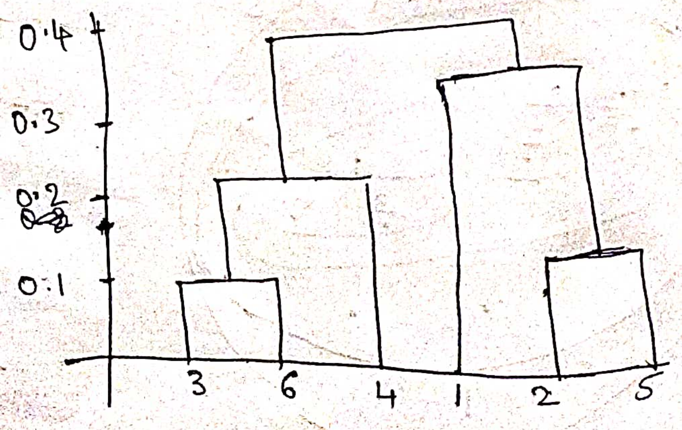
$$\text{dist}(\{3,6\},\{1\}) = \max(\text{dist}(3,1), \text{dist}(6,1))$$

$$= \max(0.22, 0.23)$$

$$= 0.23$$



Complete link clustering



Complete link dendrogram

Average Link :

Below figure shows results After Applying the group Average approach to sample dat of six points.

→ We calculate the distence between some clusters.

→ proximity ⇒ $\text{proximity}(C_i, C_j) = \dfrac{\sum_{x \in C_i, y \in C_j} \text{Proximity}(x,y)}{m_i \times m_j}$

$\text{dist}(\{3,6,4\}, \{1\}) = (0.22+0.37+0.23)/(3\times1)$

$= 0.28$

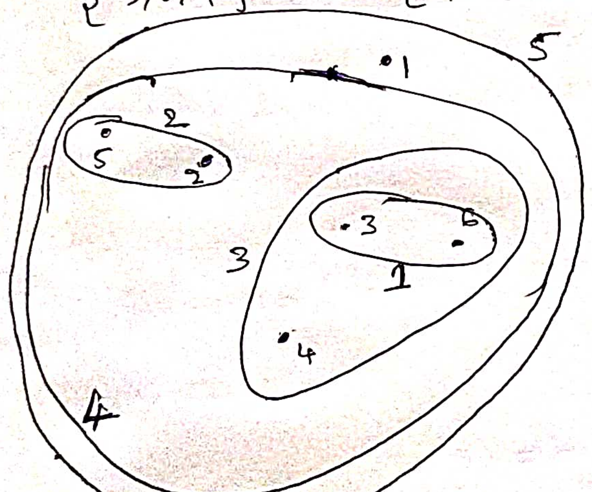
$\text{dist}(\{2,5\}, \{1\}) = (0.24+0.34)/(2\times1)$

$= 0.29$

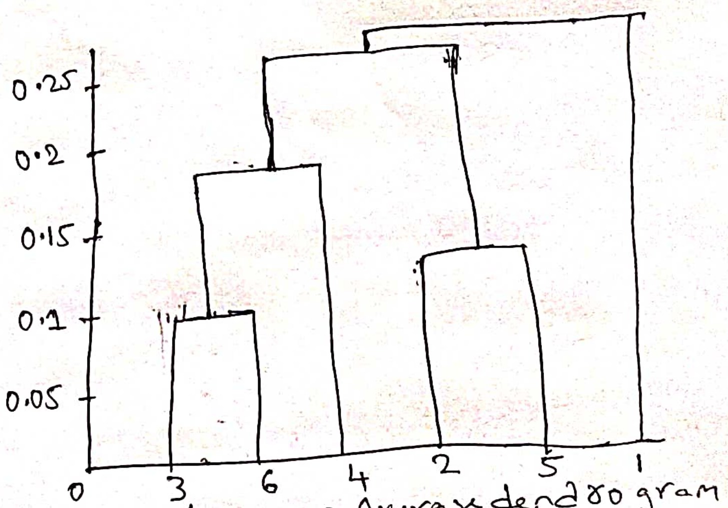
$\text{dist}(\{3,6,4\}, \{2,5\}) = (0.15+0.28+0.25+0.39+0.20+0.29)/(3\times2)$

$= 0.26$

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



Group Average clustering



(b) group Average dendrogram

→ Average version of hierarchical clustering, the proximity of two clusters is defined as the average pairwise proximity among all pairs. of points in the different clusters. proximity proximity (C_i, C_j) of clusters C_i and C_j which are of size m_i and m_j respectively is

$$\text{Proximity}(C_i, C_j) = \frac{\sum_{x \in C_i} \text{Proximity}(x,y)}{m_i \times m_j}$$

→ this is an Intermediate approach between the single and Complete link approaches.