ID - 700726283

Machine learning

Assignment - 6

(Q1) **Single Link proximity function:-** The minimum distance between any two points in the different clusters and that distance is the height at which they are joined into one cluster in the dendogram.

→ So, from the table given in the question 1

= dist $(P_3, P_6) = 0.1100$

= dist $(P_2, P_5) = 0.1388$

= dist $(\{(P_3, P_6), (P_2, P_5)\})$

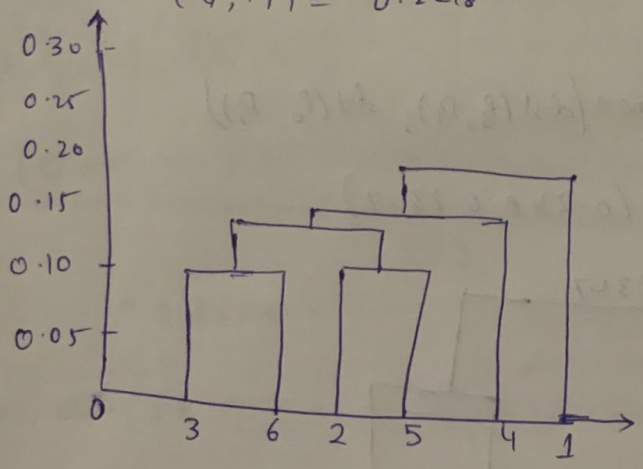
= min $(\text{dist}(P_3, P_2); \text{dist}(P_6, P_2); \text{dist}(P_3, P_5); \text{dist}(P_6, P_5))$

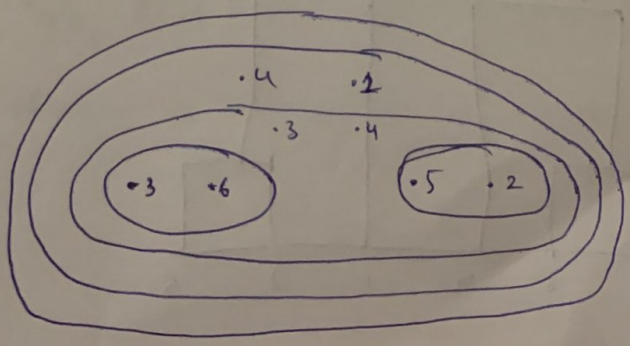
= min $(0.1483, 0.2540, 0.2593, 0.3928)$

= 0.1483

= dist $(P_3, P_4) = 0.1513$

dist $(P_4, P_1) = 0.2218$





Complete link proximity function:

In here the proximity of two clusters is defined to be the maximum of the distance b/w any two points in different cluster.

Points (P_3, P_6) are merged first and then it is merged with P_4.

$$dist(\{P_3, P_6\}, \{4\}) = \max(dist(P_3, P_4); dist(P_6, P_4))$$

$$= \max(0.1513, 0.2216)$$

$$= 0.2216.$$

$$dist(\{P_3, P_6\}, \{P_2, P_5\}) = \max(dist(P_3, P_2); dist(P_3, P_5); dist(P_6, P_2); dist(P_6, P_5))$$

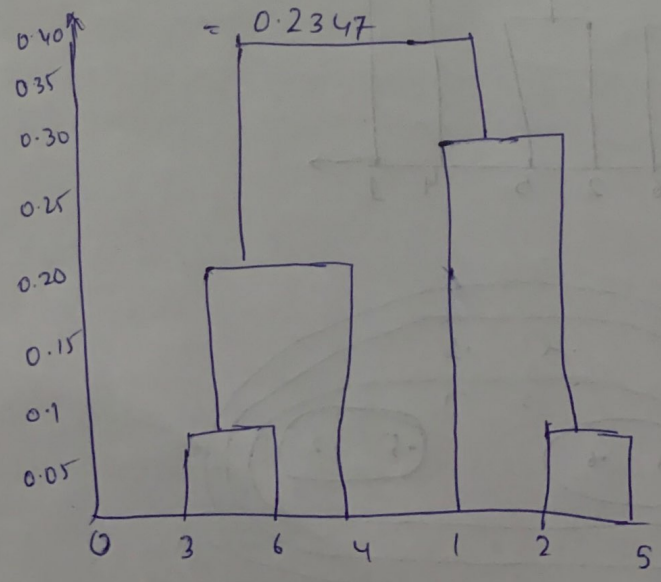
$$= \max(0.1483, 0.2540; 0.2843, 0.3921)$$

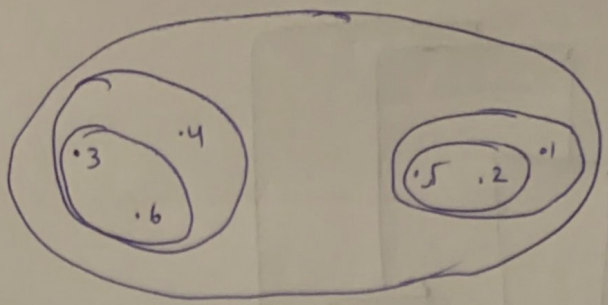
$$= 0.3921$$

$$dist(\{P_3, P_6\}, \{P_1\}) = \max(dist(P_3, P_1); dist(P_6, P_1))$$

$$= \max(0.2318, 0.2347)$$

$$= 0.2347$$





Average Link proximity function :

The average of the pairwise proximities between all pairs of points in the different clusters, This is the intermediate approach b/w single and complete Link.

$$\Rightarrow \text{proximity}(\text{cluster1}, \text{cluster2}) = \sum_{\substack{P_1 \in \text{cluster1} \\ P_2 \in \text{cluster2}}} \frac{\text{Proximity}(P_1, P_2)}{\text{size(cluster 1)} \; \text{size(cluster2)}}$$

- $\text{dist}(\{P_3, P_6, P_4\}, \{P_1\}) = (0.2218 + 0.20347 + 0.3688)/3 \times 1$

$= 0.2751$

$= \text{dist}(\{P_2, P_5\}, \{P_1\}) = (0.2357 + 0.3421)/2 \times 1$

$= 0.2889$

$\text{dist}(\{P_3, P_6, P_4\}, \{P_2, P_5\}) =$

$$= \frac{(0.1483 + 0.2843 + 0.2540 + 0.3421 + 0.2047 + 0.2932)}{6 \times 2}$$

$= 0.2637$

Because $\text{dist}(\{P_3, P_6, P_4\}, \{P_2, P_5\})$ is smaller than $\text{dist}(\{P_3, P_6, P_4\}, \{P_1\})$ and $\text{dist}(\{P_2, P_5\}, \{P_1\})$

These two clusters are merged at the fourth stage.