

Q1) Consider the 5 points (p1-p5) whose coordinates are given below which belong to a two-dimensional vector space. The dataset containing these points will be clustered into 3 clusters. The L_1 norm will be used as the "closeness" measure to determine the centroid each point is "close" to.

P1	P2	P3	P4	P5
(0,3)	(0,4)	(0,5)	(9.5,8.7)	(10,9)

A) Find the clusters each point belongs to using the **single link technique (min)**. Show your work clearly. (17 pts)

Find (fill) the proximity matrix
 \rightarrow we will start with a cluster for each point

	P1	P2	P3	P4	P5
P1	0	1	2	15.2	16
P2	1	0	1	14.2	15
P3	2	1	0	13.2	14
P4	15.2	14.2	13.2	0	0.8
P5	16	15	14	0.8	0

$$\text{dist}(P1, P2) = |0-0| + |3-4| = 1 \text{ (same as } \text{dist}(P2, P1))$$

$$\text{dist}(P2, P3) = |0-0| + |4-5| = 1$$

$$\text{dist}(P1, P4) = |0-9.5| + |3-8.7| = 15.2$$

$$\text{dist}(P1, P5) = |0-10| + |3-9| = 16$$

$$\text{dist}(P2, P3) = |0-0| + |4-5| = 1$$

$$\text{dist}(P2, P4) = |0-9.5| + |4-8.7| = 14.2$$

$$\text{dist}(P2, P5) = |0-10| + |4-9| = 15$$

$$\text{dist}(P3, P4) = |0-9.5| + |5-8.7| = 13.2$$

$$\text{dist}(P3, P5) = |0-10| + |5-9| = 14$$

$$\text{dist}(P4, P5) = |9.5-10| + |8.7-9| = 0.8$$

We will group the two most similar points (least distance)
 (P4, P5)

	P ₁	P ₂	P ₃	P ₄ U _{P₅}
P ₁	0	1	2	15.2
P ₂	1	0	1	14.2
P ₃	2	1	0	13.2
P ₄ U _{P₅}	15.2	14.2	13.2	0

$$\text{dist}(P_1, (P_4 \cup P_5)) = \min(\text{dist}(P_1, P_4), \text{dist}(P_1, P_5)) \\ = \min(15.2, 16) = 15.2$$

Note ① in case of using complete link (max):

$$\text{dist}(P_1, (P_4 \cup P_5)) = \max(\text{dist}(P_1, P_4), \text{dist}(P_1, P_5)) \\ = \max(15.2, 16) = 16$$

② in case of using average link:

$$\text{dist}(P_1, (P_4 \cup P_5)) = \frac{\text{dist}(P_1, P_4) + \text{dist}(P_1, P_5)}{2} \quad (\text{average}) \\ = \frac{15.2 + 16}{2} = 15.6$$

$$\text{dist}(P_2, (P_4 \cup P_5)) = \min(\text{dist}(P_2, P_4), \text{dist}(P_2, P_5)) \\ = \min(14.2, 15) = \underline{14.2}$$

$$\text{dist}(P_3, (P_4 \cup P_5)) = \min(\text{dist}(P_3, P_4), \text{dist}(P_3, P_5)) \\ = \min(13.2, 14) = 13.2$$

we will group (P₁, P₂) OR (P₂, P₃)

I will group (P₁, P₂)

	$P_1 \cup P_2$	P_3	$P_4 \cup P_5$
$P_1 \cup P_2$	0	1	14.2
P_3	1	0	13.2
$P_4 \cup P_5$	14.2	13.2	0

$$\text{dist}((P_1 \cup P_2), P_3) = \min(\text{dist}(P_1, P_3), \text{dist}(P_2, P_3)) \\ = \min(2, 1) = 1$$

$$\text{dist}((P_1 \cup P_2), (P_4 \cup P_5)) = \min(\text{dist}(P_1, (P_4 \cup P_5)), \text{dist}(P_2, (P_4 \cup P_5))) \\ = \min(15.2, 14.2) = 14.2$$

we will group $(P_3, (P_1 \cup P_2))$

	$(P_1 \cup P_2) \cup P_3$	$P_4 \cup P_5$
$(P_1 \cup P_2) \cup P_3$ (A)	0	13.2
$P_4 \cup P_5$ (B)	13.2	0

$$\text{dist}(A, B) = \min(\text{dist}((P_1 \cup P_2), (P_4 \cup P_5)), \text{dist}(P_3, (P_4 \cup P_5))) \\ = \min(14.2, 13.2) = 13.2$$

B) Draw the clusters dendrogram. (5 pts)

