

Clustering (Agglomerative)

(Q1) Discuss the agglomerative algorithm with the following data and plot a dendrogram using single linkage approach. The table below comprises sample data items including distances between elements

Item	E	A	C	B	D
E	0	1	2	2	3
A	1	0	2	5	3
C	2	2	0	1	6
B	2	5	1	0	3
D	3	3	6	3	0

=> Item	E	A	C	B	D
E	0	1	2	2	3
A	1	0	2	5	3
C	2	2	0	1	6
B	2	5	1	0	3
D	3	3	6	3	0

Combining A, E

Item	E, A	C	B	D
E, A	0			
C	2	0		
B	2	1	0	
D	3	6	3	0

$$\begin{aligned}
 & \min \{ \dots, (C, E), \dots, (C, A) \} \Rightarrow \{2, 2\} \Rightarrow 2 \\
 & \min \{ \dots, (B, E), \dots, (B, A) \} \Rightarrow \{2, 5\} \Rightarrow 2 \\
 & \min \{ \dots, (D, E), \dots, (D, A) \} \Rightarrow \{3, 3\} \Rightarrow 3 \\
 & \min(B, \dots) \Rightarrow 1
 \end{aligned}$$

Combining B, C

Item	E, A	B, C	D
E, A	0		
B, C	2	0	
D	3	3	0

$$\min \{d(E, B), (E, C), (A, B), (A, C)\}$$

$$\min \{2, 2, 5, 2\} \Rightarrow 2$$

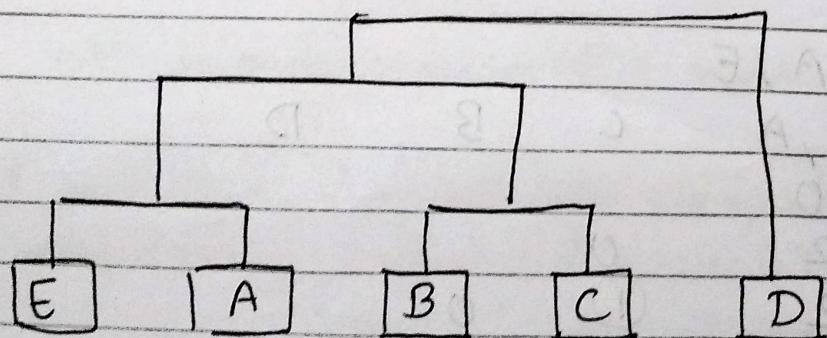
$$\min \{d(E, D), (A, D)\} \Rightarrow (3, 3) \Rightarrow 3$$

$$\min \{d(B, D), (C, D)\} \Rightarrow (3, 6) \Rightarrow 3$$

combining E, A with B, C

Item	E, A, B, C	D
E, A, B, C	0	
D	3	0

Dendogram:



(Q2)

Item	P1	P2	P3	P4	P5
P1	0				
P2	9	0			
P3	3	7	0		
P4	6	5	9	0	
P5	11	10	2	8	0

Item	P1	P2	P3	P4	P5
P1	0				
P2	9	0			
P3	3	7	0		
P4	6	5	9	0	
P5	11	10	(2)	8	0

Item	P1	P2	P3, P5	P4
P1	0			
P2	9	0		
P3, P5	(3)	7	0	
P4	6	5	8	0

$$\min \{ (P_1, P_3), (P_1, P_5) \} \Rightarrow \min(3, 11) \Rightarrow 3$$

$$\min \{ (P_2, P_3), (P_2, P_5) \} \Rightarrow \min(7, 10) \Rightarrow 7$$

$$\min \{ (P_3, P_4), (P_5, P_4) \} \Rightarrow \min(9, 8) \Rightarrow 8$$

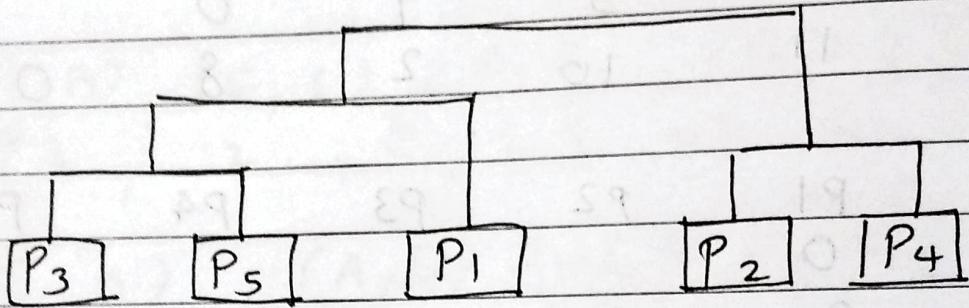
Item	P1, P3, P5	P2	P4
P1, P3, P5	0		
P2	7	0	
P4	6	(5)	0

$$\min \{ (P_1, P_2), (P_3, P_2), (P_5, P_2) \} \Rightarrow \{ 9, 7, 10 \} \Rightarrow 7$$

$$\min \{ (P_1, P_4), (P_3, P_4), (P_5, P_4) \} \Rightarrow \{ 6, 9, 8 \} \Rightarrow 6$$

Item	P1, P3, P5	P2, P4
P1, P3, P5	0	0
P2, P4	6	0

Dendrogram:



C = {(1, 2) since $\{(1, 2)\} \subset \{(1, 2), (3, 4), (5, 6)\}$ }
 E = {(1, 2) since $\{(1, 2)\} \subset \{(1, 2), (3, 4), (5, 6)\}$ }
 F = { } since $\{ \} \subset \{(1, 2), (3, 4), (5, 6)\}$

Q3) Following table gives fat & protein content of items. Apply single linkage clustering and construct dendrogram

Food item	Protein	Fat
1	1.1	60
2	8.2	20
3	4.2	35
4	1.5	21
5	7.6	15
6	2.0	55
7	3.9	39

$$\Rightarrow d(1, 2) = \sqrt{(8.2 - 1.1)^2 + (20 - 60)^2} \\ = 40.6$$

$$d(1, 3) = \sqrt{(4.2 - 1.1)^2 + (35 - 60)^2} \\ = 25.2$$

$$d(1, 4) = \sqrt{(1.5 - 1.1)^2 + (21 - 60)^2} \\ = 39$$

$$d(1, 5) = \sqrt{(7.6 - 1.1)^2 + (15 - 60)^2} \\ = 45.5$$

$$d(1, 6) = \sqrt{(2.0 - 1.1)^2 + (55 - 60)^2} \\ = 5.1$$

$$d(1, 7) = \sqrt{(3.9 - 1.1)^2 + (39 - 60)^2} \\ = 21.2$$

$$d(2, 3) = \sqrt{(4.2 - 8.2)^2 + (35 - 20)^2} \\ = 15.5$$

$$d(2,4) = \sqrt{(1.5 - 8.2)^2 + (21 - 20)^2} \\ = 6.8$$

$$d(2,5) = \sqrt{(7.6 - 8.2)^2 + (15 - 20)^2} \\ = 5.03$$

$$d(2,6) = \sqrt{(2.0 - 8.2)^2 + (55 - 20)^2} \\ = 35.5$$

$$d(2,7) = \sqrt{(3.9 - 8.2)^2 + (39 - 20)^2} \\ = 19.5$$

$$d(3,4) = \sqrt{(1.5 - 4.2)^2 + (21 - 35)^2} \\ = 14.3$$

$$d(3,5) = \sqrt{(7.6 - 4.2)^2 + (15 - 35)^2} \\ = 20.9$$

$$d(3,6) = \sqrt{(2.0 - 4.2)^2 + (55 - 35)^2} \\ = 20.1$$

$$d(3,7) = \sqrt{(3.9 - 4.2)^2 + (39 - 35)^2} \\ = 4.01$$

$$d(4,5) = \sqrt{(7.6 - 1.5)^2 + (15 - 21)^2} \\ = 8.6$$

$$d(4,6) = \sqrt{(2.0 - 1.5)^2 + (55 - 21)^2} \\ = 34$$

$$d(4,7) = \sqrt{(3.9 - 1.5)^2 + (39 - 21)^2} \\ = 18.2$$

→ Here I should have chose 4.01, a minor mistake
 but it doesn't change the final answer.
 Final answer is correct.

$$d(5,6) = \sqrt{(2.0 - 7.6)^2 + (55 - 15)^2} \\ = 46.4$$

$$d(5,7) = \sqrt{(3.9 - 7.6)^2 + (39 - 15)^2} \\ = 24.3$$

$$d(6,7) = \sqrt{(3.9 - 2.0)^2 + (39 - 55)^2} \\ = 16.1$$

Item	1	2	3	4	5	6	7
1	0					0	
2	40.6	0					
3	25.2	15.5	0				
4	39	6.8	14.3	0			
5	45.5	5.03	20.9	8.6	0		
6	5.1	35.5	20.1	34	40.4	0	
7	21.2	19.5	4.01	18.2	24.3	16.1	0

Combining 2,5

Item	1	2,5	3	4	6	7
1	0					
2,5	40.6	0				
3	25.2	20.9	0			
4	39	6.8	14.3	0		
6	5.1	35.5	20.1	34	0	
7	21.2	19.5	4.01	18.2	16.1	0

$$\min \{ (1,2), (1,5) \} \Rightarrow 40.6$$

$$\min \{ (3,2), (3,5) \} \Rightarrow 20.9$$

$$\min \{ (4,2), (4,5) \} \Rightarrow 6.8$$

$$\min \{ (6,2), (6,5) \} \Rightarrow 35.5$$

$$\min \{ (7,2), (7,5) \} \Rightarrow 19.5$$

Combining 1, 6

Items	1, 6	2, 5	3	4	7
1, 6	0				
2, 5	35.5	0			
3	20.1	20.9	0		
4	34	6.8	14.3	0	
7	16.1	19.5	4.01	18.2	0

Combining 2, 5 with 4

Items	1, 6	2, 5, 4	3	7
1, 6	0			
2, 5, 4	34	0		
3	20.1	14.3	0	
7	16.1	18.2	4.01	0

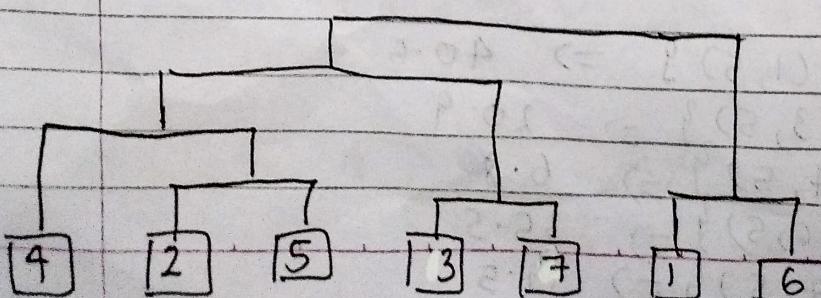
Combining 3 with 7

Items	1, 6	2, 5, 4	3, 7
1, 6	0		
2, 5, 4	34	0	
3, 7	16.1	14.3	0

Combining 2, 5, 4 with 3, 7

Items	1, 6	2, 3, 4, 5, 7
1, 6	0	
2, 3, 4, 5, 7	16.1	0

Dendrogram :



(Q4) Find the clusters for the following dataset using a single-link technique. Use euclidean distance and draw the dendrogram

Sample no	X	Y
P1	0.40	0.53
P2	0.22	0.38
P3	0.35	0.32
P4	0.26	0.19
P5	0.08	0.41
P6	0.45	0.30

$$\Rightarrow d(P_1, P_2) = \sqrt{(0.22 - 0.40)^2 + (0.38 - 0.53)^2} \\ = 0.23$$

$$d(P_1, P_3) = \sqrt{(0.35 - 0.40)^2 + (0.32 - 0.53)^2} \\ = 0.22$$

$$d(P_1, P_4) = \sqrt{(0.26 - 0.40)^2 + (0.19 - 0.53)^2} \\ = 0.37$$

$$d(P_1, P_5) = \sqrt{(0.08 - 0.40)^2 + (0.41 - 0.53)^2} \\ = 0.34$$

$$d(P_1, P_6) = \sqrt{(0.45 - 0.40)^2 + (0.30 - 0.53)^2} \\ = 0.24$$

$$d(P_2, P_3) = \sqrt{(0.35 - 0.22)^2 + (0.32 - 0.38)^2} \\ = 0.14$$

$$d(P_2, P_4) = \sqrt{(0.26 - 0.22)^2 + (0.19 - 0.38)^2} \\ = 0.19$$

$$d(P_2, P_5) = \sqrt{(0.08 - 0.22)^2 + (0.41 - 0.38)^2} \\ = 0.14$$

$$d(P_2, P_6) = \sqrt{(0.45 - 0.22)^2 + (0.30 - 0.38)^2} \\ = 0.24$$

$$d(P_3, P_4) = \sqrt{(0.26 - 0.35)^2 + (0.19 - 0.32)^2} \\ = 0.13$$

$$d(P_3, P_5) = \sqrt{(0.08 - 0.35)^2 + (0.41 - 0.32)^2} \\ = 0.28$$

$$d(P_3, P_6) = \sqrt{(0.45 - 0.35)^2 + (0.30 - 0.32)^2} \\ = 0.10$$

$$d(P_4, P_5) = \sqrt{(0.08 - 0.26)^2 + (0.41 - 0.19)^2} \\ = 0.23$$

$$d(P_4, P_6) = \sqrt{(0.45 - 0.26)^2 + (0.30 - 0.19)^2} \\ = 0.22$$

$$d(P_5, P_6) = \sqrt{(0.45 - 0.08)^2 + (0.30 - 0.41)^2} \\ = 0.39$$

Items	P1	P2	P3	P4	P5	P6
P1	0					
P2	0.23	0				
P3	0.22	0.14	0.10	0.13	0	
P4	0.37	0.19	0.13	0		
P5	0.34	0.14	0.28	0.23	0	
P6	0.24	0.24	0.10	0.22	0.39	0

Combining P₃ with P₆

Items	P ₁	P ₂	P _{3, P₆}	P ₄	P ₅
P ₁	0				
P ₂	0.23	0			
P _{3, P₆}	0.22	0.14	0		
P ₄	0.37	0.19	0.13	0	
P ₅	0.34	0.14	0.28	0.23	0

$$\min \{ (P_1, P_3), (P_1, P_6) \} \Rightarrow 0.22$$

$$\min \{ (P_2, P_3), (P_2, P_6) \} \Rightarrow 0.14$$

$$\min \{ (P_4, P_3), (P_4, P_6) \} \Rightarrow 0.13$$

$$\min \{ (P_5, P_3), (P_5, P_6) \} \Rightarrow 0.28$$

∴ Combining P_{3, P₆} with P₄

Items	P ₁	P ₂	P _{3, P₄, P₆}	P ₅
P ₁	0			
P ₂	0.23	0		
P _{3, P₄, P₆}	0.22	0.14	0	
P ₅	0.34	0.14	0.28	0

Combining P_{2, P₅} with P₆

Items	P ₁	P _{2, P₅}	P _{3, P₄, P₆}
P ₁	0		
P _{2, P₅}	0.23	0	
P _{3, P₄, P₆}	0.22	0.14	0

Combining P_{2, P₅} with P_{3, P₄, P₆}

Items	P ₁	P _{2, P₃, P₄, P₅, P₆}
P ₁	0	
P _{2, P₃, P₄, P₅, P₆}	0.22	0

Dendogram

