HomeWork-4

Part-A : Calculate

1) Step-1 : Compute Euclidean Distance Matrix.

$$d(P_i, P_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

From/To	P_1	P_2	P_3	P_4	P_5	P_6	P_7
P_1	0	0.283	0.672	0.628	0.372	0.254	0.665
P_2	0.283	0	0.231	0.182	0.412	0.141	0.190
P_3	0.672	0.231	0	0.111	0.505	0.264	0.072
P_4	0.628	0.182	0.111	0	0.393	0.134	0.104
P_5	0.372	0.412	0.505	0.393	0	0.315	0.496
P_6	0.254	0.141	0.264	0.134	0.135	0	0.230
P_7	0.665	0.190	0.072	0.104	0.496	0.230	0

Step-2 : Single Linkage (MIN) Method

i) min = P3-P7 = 0.072

→ merge (P_3, P_7)

New cluster $C_1 = (P_3, P_7)$

2) min distances b/w C_1 & other using MIN scale.

$d(C_1, P_4) = \min(d(P_3, P_4), d(P_7, P_4)) = \min(0.111, 0.104)$

$= 0.104$

→ merge $(C_1, P_4) = (P_3, P_4, P_7)$

New cluster $C_2 = (P_3, P_4, P_7)$

3) $d(C_2, P_2) = \min(0.231, 0.182, 0.192) = 0.170$

→ merge $(P_2, P_3) = \{P_2, P_3, P_4, P_7\}$

4) $d(C_5, P_6) = \min(0.256, 0.134, 0.230, 0.241) = 0.134$

merge $(C_5, P_6) = \{P_2, P_3, P_4, P_6, P_7\}$

5) $d(C_2, P_1) = \min(0.283, 0.492, 0.98, 0.256, 0.465) = 0.256$

merge (C_2, P_1)

6) Finally

$d(C_2, P_5) = \min(0.372, 0.412, 0.505, 0.393, 0.915, 0.496)$

$= 0.315$

→ merge all → single cluster

Merging order (with distances)

P_3-P_7 (0.72)

$(P_3 P_7)$ - P_4 (0.104)

+ P_6 (0.134)

+ P_2 (0.170)

+ P_1 (0.254)

+ P_5 (0.315)

Step-3: Average Linkage Method

we take the average distance b/w points of the clusters

1) min = P_3-P_7 = 0.072

cluster $C_1 = (P_3, P_7)$

2) compute avg distance blw c_1 & P_4

$d_{avg}(c_1, P_4) = (0.111 + 0.104)/2 = 0.1075$

min → merge c_1 & P_4

→ $c_2 = (P_3, P_4, P_7)$

3) Avg distance c_2-P_6

$d_{avg}(c_2, P_6) = (0.264 + 0.134 + 0.230)/3 = 0.209$

c_2-P_2: $(0.231 + 0.182 + 0.170)/3 = 0.194$.

minimum is 0.194 → merge with P_2

$c_3 = \{P_2, P_3, P_4, P_7\}$

4) c_3-P_6:

$(0.141 + 0.264 + 0.134 + 0.230)/4 = 0.192$

→ merge c_3 with P_6 (0.192)

5) $c_4 = (P_2, P_3, P_4, P_6, P_7)$

c_4-P_1:

$(0.283 + 0.472 + 0.428 + 0.254 + 0.465)/5 = 0.380$

c_4-P_5:

$(0.412 + 0.505 + 0.393 + 0.315 + 0.496)/5$

$= 0.424$

Next merge is with P_1 at 0.380

6) Finally merge C5 with P_5:

$(0.372 + 0.612 + 0.505 + 0.393 + 0.315 + 0.696)/6$

$= 0.616.$

Final Dendrogram (Average Linkage)

Merging Order (with avg distance)

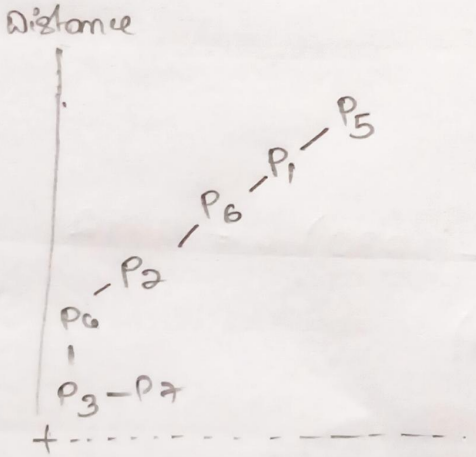
P_3-P_7 (0.072) + P_6 (0.192)

(P_3,P_7)-P_4 (0.108) + P_1 (0.380)

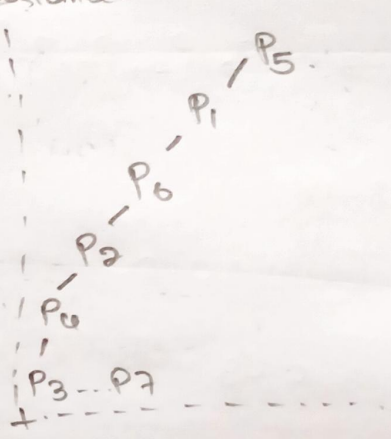
+ P_2 (0.194) + P_5 (0.616)

Step-4: Dendrogram Sketch (Summary)

Single Linkage (MIN)



Average Linkage



0.696)/6

2) Given data points

Point	X	Y
P_1	2	1
P_2	3	1
P_3	3	3
P_4	4	1
P_5	5	1
P_6	6	4
P_7	1	3
P_8	2	5

Centroids

cluster	coordinates
c_1	(2,1)
c_2	(4,1)
c_3	(5,1)

Euclidean Distance Formula

$$d = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

Step-1:-

Compute Distance of Each Point from Each centroid

Point	To $c_1(2,1)$	To $c_2(4,1)$	To $c_3(5,1)$
$P_1(2,1)$	$\sqrt{(2-2)^2+(1-1)^2}=0$	$\sqrt{(2-4)^2+(1-1)^2}$ $=2.000$	$\sqrt{(2-5)^2+(1-1)^2}$ $=3.000$
$P_2(3,1)$	$\sqrt{(3-2)^2+(1-1)^2}$ $=1.000$	$\sqrt{(3-4)^2+(1-1)^2}$ $=1.000$	$\sqrt{(3-5)^2+(1-1)^2}$ $=2.000$
$P_3(3,3)$	$\sqrt{(3-2)^2+(3-1)^2}$ $=2.236$	$\sqrt{(3-4)^2+(3-1)^2}$ $=2.236$	$\sqrt{(3-5)^2+(3-1)^2}$ $=2.828$
$P_4(4,1)$	$\sqrt{(4-2)^2+(1-1)^2}$ $=2.00$	$\sqrt{(4-4)^2+(1-1)^2}$ $=0.00$	$\sqrt{(4-5)^2+(1-1)^2}$ $=1.000$
$P_5(5,1)$	$\sqrt{(5-2)^2+(1-1)^2}$ $=3.0$	$\sqrt{(5-4)^2+(1-1)^2}$ $=1.000$	$\sqrt{(5-5)^2+(1-1)^2}$ $=0.00$
$P_6(6,2)$	$\sqrt{(6-2)^2+(2-1)^2}$ $=2.211$	$\sqrt{(6-4)^2+(2-1)^2}$ $=6.325$	$\sqrt{(6-5)^2+(2-1)^2}$ $=6.083$
$P_7(1,3)$	$\sqrt{(1-2)^2+(3-1)^2}$ $=2.236$	$\sqrt{(1-4)^2+(3-1)^2}$ $=3.606$	$\sqrt{(1-5)^2+(3-1)^2}$ $=4.472$
$P_8(2,5)$	$\sqrt{(2-2)^2+(5-1)^2}$ $=4.000$	$\sqrt{(2-4)^2+(5-1)^2}$ $=4.472$	$\sqrt{(2-5)^2+(5-1)^2}$ $=5.00$

centroid

,1)

$(1-1)^2$
3.000

$(1-1)^2$
2.000

$+(3-1)^2$
28

$(1-1)^2$
000

$(1+1)^2$
0

$+(7-1)^2$
083

$(3-1)^2$
2.672

$+(5-1)^2$
0

Step-2: Assign points to Nearest Centroid.

Point	c_1	c_2	c_3	Nearest cluster
P_1	0.000	2.000	3.000	c_1
P_2	1.000	1.000	2.000	c_1 or c_2
P_3	2.236	2.236	2.828	c_1 (tie-c_1)
P_4	2.000	0.000	1.000	c_2
P_5	3.000	1.000	0.000	c_3
P_6	7.211	6.325	6.083	c_3
P_7	2.236	3.606	4.472	c_1
P_8	4.000	3.472	5.000	c_1

Cluster Assignments

cluster	Points	coordinates
c_1	P_1, P_2, P_3, P_7 P_8	(2,1), (3,1) (3,3) (1,3) (2, 5)
c_2	P_4	(4,1)
c_3	P_5, P_6	(5,1) (6,7)

Step-3: Compute New Centroids

c_1 new centroid.

$\bar{x} = (2+3+3+1+2)/5 = 11/5 = 2.2$

$\bar{y} = (1+1+3+3+5)/5 = 13/5 = 2.6$

→ c_1-new = (2.2, 2.6)

c_2 new centroid.

only one point → (4,1)

c_3 new centroid.

$\bar{x} = (5+6)/2 = 5.5$ $\bar{y} = (1+7)/2 = 4.$

c_3-new = (5.5, 4)

* Updated Centroids after 1st iteration

cluster	New Centroid
c_1	(2.2, 2.6)
c_2	(4,1)
c_3	(5.5, 4)