MUHAMMAD MUAAZ SHOAIB
FA20-BCS-074
MACHINE LEARNING
22-MAY-2023



Problem 1 :- Let's assume we have a dataset of 8 two dimensional data points as follows:-
(2,10), (2,5), (8,4), (5,8), (7,5), (6,4), (1,2), (4,9)
Let assume we choose K=2
cluster center 1: (2,10)
cluster center 2: (5,8)

Sol:- Initial centers
C1:(2,10)
C2:(5,8)

| Data Points $(x_1, y_1)$ | C1 $(x_2, y_2)$ (2,10) | C2 $(x_2, y_2)$ (5,8) | Cluster | New Cluster |
|---|---|---|---|---|
| (2,10) | 0 | 3.61 | 1 | |
| (2,5) | 5 | 4.24 | 2 | |
| (8,4) | 8.49 | 5 | 2 | |
| (5,8) | 3.61 | 0 | 2 | |
| (7,5) | 7.07 | 3.61 | 2 | |
| (6,4) | 7.21 | 4.12 | 2 | |
| (1,2) | 8.06 | 7.21 | 2 | |
| (4,9) | 2.24 | 1.41 | 2 | |

I have calculated the distances using this formula

$$d(p_1, p_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

After calculating the distances, I have chec which value is less.

Calculating new centers

$C_1 : (2, 10)$

$C_2 : \left(\frac{2+8+5+7+6+1+4}{7}, \frac{5+4+8+5+4+2+9}{7}\right)$

$C_2 : (4.71, 5.28)$

Now calculate distances, and check which value is less

$d(P_1, P_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$

| Data Points $(x_1, y_1)$ | $C_1$ $(x_2, y_2)$ $(2, 10)$ | $C_2$ $(x_2, y_2)$ $(4.71, 5.28)$ | Cluster | New Cluster |
|---|---|---|---|---|
| (2, 10) | 0 | 5.44 | 1 | 1 |
| (2, 5) | 5 | 2.72 | 2 | 2 |
| (8, 4) | 8.49 | 3.53 | 2 | 2 |
| (5, 8) | 3.61 | 2.74 | 2 | 2 |
| (7, 5) | 7.07 | 2.31 | 2 | 2 |
| (6, 4) | 7.21 | 1.82 | 2 | 2 |
| (1, 2) | 8.06 | 4.95 | 2 | 2 |
| (4, 9) | 2.27 | 3.78 | 2 | 1 |

Calculating new centers

$C_1 : \left(\frac{2+4}{2}, \frac{10+9}{2}\right)$

$C_1 : (3, 9.5)$

$$: \left( \frac{2+8+5+7+6+1}{6}, \frac{5+4+8+5+4+2}{6} \right)$$

$C_2 : (4.83, 4.66)$

Now calculate distances, check which value is less.

| Data Points $(x_1, y_1)$ | $C_1$ $(x_2, y_2)$ $(3, 9.5)$ | $C_2$ $(x_2, y_2)$ $(4.83, 4.66)$ | Cluster | New Cluster |
|---|---|---|---|---|
| (2, 10) | 1.12 | 6.04 | 1 | 1 |
| (2, 5) | 4.61 | 2.85 | 2 | 2 |
| (8, 4) | 7.43 | 3.24 | 2 | 2 |
| (5, 8) | 2.5 | 3.34 | 2 | 1 |
| (7, 5) | 6.02 | 2.19 | 2 | 2 |
| (6, 4) | 6.26 | 1.34 | 2 | 2 |
| (1, 2) | 7.76 | 4.66 | 2 | 2 |
| (4, 9) | 1.12 | 4.41 | 1 | 1 |

Calculating new centers

$$C_1 : \left( \frac{2+5+4}{3}, \frac{10+8+9}{3} \right)$$

$C_1 : (3.66, 9)$

$$C_2 : \left( \frac{2+8+7+6+1}{5}, \frac{5+4+5+4+2}{5} \right)$$

$C_2 : (4.8, 4)$

Now calculate distances, and check which value is less

$$d(P_1, P_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

| Data Points $(x_1, y_1)$ | $C_1$ $(x_2, y_2)$ (3.66, 9) | $C_2$ $(x_2, y_2)$ (4.8, 4) | Cluster | New Cluster |
|---|---|---|---|---|
| (2, 10) | 1.93 | 6.62 | 1 | 1 |
| (2, 5) | 4.33 | 2.97 | 2 | 2 |
| (8, 4) | 6.62 | 3.2 | 2 | 2 |
| (5, 8) | 1.67 | 4.00 | 1 | 1 |
| (7, 5) | 5.21 | 2.41 | 2 | 2 |
| (6, 4) | 5.52 | 1.2 | 2 | 2 |
| (1, 2) | 7.48 | 4.29 | 2 | 2 |
| (4, 9) | 0.34 | 5.06 | 1 | 1 |

Hence we achieved convergence

→ Points belong to cluster 1 are
(2, 10), (5, 8), (4, 9)

→ Points belong to cluster 2 are
(2, 5), (8, 4), (7, 5), (6, 4), (1, 2)

——— x ——— x ———

Problem 2:- Dry run for hierarchical clustering using the agglomerative (bottom up) approach for the given data as under:

1. (2,10)
2. (2,5)
3. (8,4)
4. (5,8)
5. (7,5)
6. (6,4)
7. (1,2)
8. (4,9)

Sol:-

$P_1(2,10)$
$P_2(2,5)$
$P_3(8,4)$
$P_4(5,8)$
$P_5(7,5)$
$P_6(6,4)$
$P_7(1,2)$
$P_8(4,9)$

Step 1:- Compute the distance matrix

$d(P_1,P_2) = \sqrt{(2-2)^2 + (5-10)^2} = 5$

$d(P_1,P_3) = \sqrt{(8-2)^2 + (4-10)^2} = 8.48$

$d(P_1,P_4) = \sqrt{(5-2)^2 + (8-10)^2} = 3.61$

$d(P_1,P_5) = \sqrt{(7-2)^2 + (5-10)^2} = 7.07$

$d(P_1,P_6) = \sqrt{(6-2)^2 + (4-10)^2} = 7.21$

$d(P_1,P_7) = \sqrt{(1-2)^2 + (2-10)^2} = 8.06$

$d(P_1,P_8) = \sqrt{(4-2)^2 + (9-10)^2} = 2.24$

$d(P_2, P_3) = \sqrt{(8-2)^2 + (4-5)^2} = 6.08$

$d(P_2, P_4) = \sqrt{(5-2)^2 + (8-5)^2} = 4.24$

$d(P_2, P_5) = \sqrt{(7-2)^2 + (5-5)^2} = 5$

$d(P_2, P_6) = \sqrt{(6-2)^2 + (4-5)^2} = 4.123$

$d(P_2, P_7) = \sqrt{(1-2)^2 + (2-5)^2} = 3.16$

$d(P_2, P_8) = \sqrt{(4-2)^2 + (9-5)^2} = 4.47$

$d(P_3, P_4) = \sqrt{(5-8)^2 + (8-4)^2} = 5$

$d(P_3, P_5) = \sqrt{(7-8)^2 + (5-4)^2} = 1.41$

$d(P_3, P_6) = \sqrt{(6-8)^2 + (4-4)^2} = 2$

$d(P_3, P_7) = \sqrt{(1-8)^2 + (2-4)^2} = 7.28$

$d(P_3, P_8) = \sqrt{(4-8)^2 + (9-4)^2} = 6.40$

$d(P_4, P_5) = \sqrt{(7-5)^2 + (5-8)^2} = 3.61$

$d(P_4, P_6) = \sqrt{(6-5)^2 + (4-8)^2} = 4.12$

$d(P_4, P_7) = \sqrt{(1-5)^2 + (2-8)^2} = 7.21$

$d(P_4, P_8) = \sqrt{(4-5)^2 + (9-8)^2} = 1.41$

$d(P_5, P_6) = \sqrt{(6-7)^2 + (4-5)^2} = 1.41$

$d(P_5, P_7) = \sqrt{(1-7)^2 + (2-5)^2} = 6.71$

$d(P_5, P_8) = \sqrt{(4-7)^2 + (9-5)^2} = 5$

$d(P_6, P_7) = \sqrt{(1-6)^2 + (2-4)^2} = 5.38$

$d(P_6, P_8) = \sqrt{(4-6)^2 + (9-4)^2} = 5.38$

$d(P_7, P_8) = \sqrt{(4-1)^2 + (9-2)^2} = 7.62$

| | $P_1$ | $P_2$ | $P_3$ | $P_4$ | $P_5$ | $P_6$ | $P_7$ | $P_8$ |
|---|---|---|---|---|---|---|---|---|
| $P_1$ | 0 | | | | | | | |
| $P_2$ | 5 | 0 | | | | | | |
| $P_3$ | 8.48 | 6.08 | 0 | | | | | |
| $P_4$ | 3.61 | 4.24 | 5 | 0 | | | | |
| $P_5$ | 7.07 | 5 | (1.41) | 3.61 | 0 | | | |
| $P_6$ | 7.21 | 4.123 | 2 | 4.12 | (1.41) | 0 | | |
| $P_7$ | 8.06 | 3.16 | 7.28 | 7.21 | 6.71 | 5.38 | 0 | |
| $P_8$ | 2.24 | 4.47 | 6.40 | (1.41) | 5 | 5.38 | 7.62 | 0 |

Step 2 :- Merging the two closest members and update distance matrix.

**Merge $P_4$ and $P_8$**

| | $P_1$ | $P_2$ | $P_3$ | $P_4, P_8$ | $P_5$ | $P_6$ | $P_7$ |
|---|---|---|---|---|---|---|---|
| $P_1$ | 0 | | | | | | |
| $P_2$ | 5 | 0 | | | | | |
| $P_3$ | 8.48 | 6.08 | 0 | | | | |
| $P_4, P_8$ | 3.61 | 4.24 | 5 | 0 | | | |
| $P_5$ | 7.07 | 5 | (1.41) | 3.61 | 0 | | |
| $P_6$ | 7.21 | 4.123 | 2 | 4.12 | (1.41) | 0 | |
| $P_7$ | 8.06 | 3.16 | 7.28 | 7.21 | 6.71 | 5.38 | 0 |

$(P_4, P_8)$

## Merge $P_5$ and $P_6$

| | $P_1$ | $P_2$ | $P_3$ | $P_4, P_8$ | $P_5, P_6$ | $P_7$ |
|---|---|---|---|---|---|---|
| $P_1$ | 0 | | | | | |
| $P_2$ | 5 | 0 | | | | |
| $P_3$ | 8.48 | 6.08 | 0 | | | |
| $P_4, P_8$ | 3.61 | 4.24 | 5 | 0 | | |
| $P_5, P_6$ | 7.07 | 5 | (1.41) | 3.61 | 0 | |
| $P_7$ | 8.06 | 3.16 | 7.28 | 7.21 | 6.71 | 0 |

$(P_4, P_8)$ $(P_5, P_6)$

## Merge $P_3$ and $(P_5, P_6)$

| | $P_1$ | $P_2$ | $P_3, (P_5, P_6)$ | $P_4, P_8$ | $P_7$ |
|---|---|---|---|---|---|
| $P_1$ | 0 | | | | |
| $P_2$ | 5 | 0 | | | |
| $P_3, (P_5, P_6)$ | 8.48 | 6.08 | 0 | | |
| $P_4, P_8$ | 3.61 | 4.24 | 5 | 0 | |
| $P_7$ | 8.06 | (3.16) | 7.28 | 7.21 | 0 |

$(P_4, P_8)$ $(P_3, (P_5, P_6))$

ge $P_2$ and $P_7$

| | $P_1$ | $P_2, P_7$ | $P_3, (P_5, P_6)$ | $P_4, P_8$ |
|---|---|---|---|---|
| 1 | 0 | | | |
| $P_2, P_7$ | 5 | 0 | | |
| $P_3, (P_5, P_6)$ | 8.48 | 6.08 | 0 | |
| $P_4, P_8$ | (3.61) | 4.24 | 5 | 0 |

$(P_2, P_7)$ $((P_3, (P_5, P_6)), (P_4, P_8))$

Merge $P_1$ and $(P_4, P_8)$

| | $P_1, (P_4, P_8)$ | $(P_2, P_7)$ | $P_3, (P_5, P_6)$ |
|---|---|---|---|
| $P_1, (P_4, P_8)$ | 0 | | |
| $(P_2, P_7)$ | (5) | 0 | |
| $P_3, (P_5, P_6)$ | 8.48 | 6.08 | 0 |

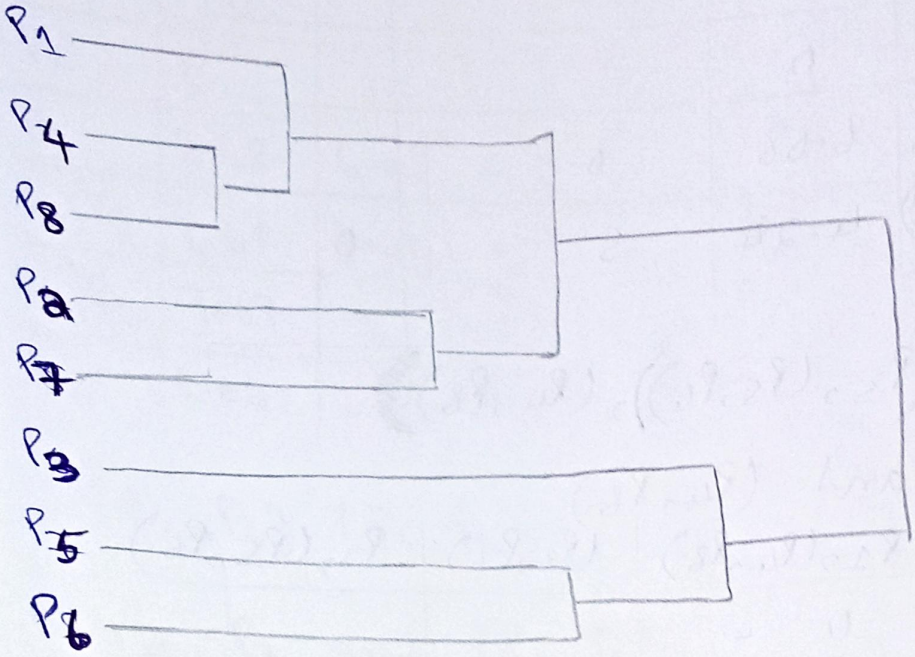
$(P_1, (P_4, P_8)), (P_3, (P_5, P_6)), (P_2, P_7)$

Merge $(P_2, P_7)$ and $(P_1, (P_4, P_8))$

| | $(P_1, (P_4, P_8)), (P_2, P_7)$ | $P_3, (P_5, P_6)$ |
|---|---|---|
| $(P_1, (P_4, P_8)), (P_2, P_7)$ | 0 | |
| $P_3, (P_5, P_6)$ | (8.48) | 0 |

$((P_1, (P_4, P_8)), (P_2, P_7)), (P_3, (P_5, P_6))$

$((P_1, (P_4, P_8)), (P_2, P_7)), (P_3, (P_5, P_6))$




Dendrogram of the cluster formed.

———— x ———— x ————