M T W T F S S
Page No.:
Date: 04-02-2025
YOUVA

Assignment 2 (K-mean clustering)

Question:

Cluster the following eight points (with (x,y) representing locations) into three clusters:

A1 (2, 10),
A2 (2, 5),
A3 (8, 4),
A4 (5, 8),
A5 (7, 5),
A6 (6, 4),
A7 (1, 2),
A8 (4, 9).

Initial clusters are:

A1 (2, 10),
A4 (5, 8). and
A7 (1, 2).

The distance function between two points a = (x_1, y_1) and b = (x_2, y_2) is defined as

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

Solution:

The initial cluster centers are:

c1 (2, 10)
c2 (5, 8)
c3 (1, 2)

Distance calculations:

point	Co-ordinates	Distance to c1(2,10)	Distance to c2(5,6)	Distance to c3(1,2)	Nearest cluster
A1	(2, 10)	0	3.61	8.06	c1
A2	(2, 5)	5	4.24	3.16	c3
A3	(8, 4)	8.49	5.00	7.28	c2
A4	(5, 8)	3.61	0	7.21	c2
A5	(7, 5)	7.07	3.61	6.71	c2
A6	(6, 4)	7.21	4.12	5.39	c2
A7	(1, 2)	8.06	7.21	0	c3
A8	(4, 9)	2.24	1.41	7.07	c2

c1 : (2,10).
c2 : A3, A4, A5, A6, A8.
c3 : A2, A7.

finding means

↳ c1 (2, 10).

↳ c2 :

mean of x-coordinates:

$$\frac{8+5+7+6+4}{5} = 6.$$

mean of y-coordinates:

$$\frac{4+8+5+4+9}{5} = 6.$$

c2 (6, 6).

↳ C3

mean of x-coordinates.

$$\frac{2+1}{2} = 1.5.$$

mean of y-coordinates.

$$\frac{5+2}{2} = 3.5.$$

C3 (1.5, 3.5).

Distance calculation:

Point	Coordinates	Distance to C1 (2,10)	Distance to C2 (6,6)	Distance to C3 (1.5, 3.5)	Nearest Cluster
A1	(2,10)	0	5.66	6.50	C1
A2	(2,5)	5	4.12	1.58	C3
A3	(8,4)	8.49	2.83	6.80	C2
A4	(5,8)	3.61	2.24	5.41	C2
A5	(7,5)	7.07	1.41	5.70	C2
A6	(6,4)	7.21	2.00	4.61	C2
A7	(1,2)	8.06	6.40	1.58	C3
A8	(4,9)	2.24	3.61	5.70	C1

update cluster centers:

2) C1 : A1, A8.

mean of x-coordinates:

$$\frac{2+4}{2} = 3.$$

mean of y-coordinates.

$$\frac{10+9}{2} = 9.5.$$

$C_1(3, 9.5)$.

⇒ C_2 : A3, A4, A5, A6.

mean of x-coordinates:

$$\frac{8+5+7+6}{4} = 6.5.$$

mean of y-coordinates:

$$\frac{4+8+5+4}{4} = 5.25.$$

$C_2(6.5, 5.25)$.

⇒ C_3 : A2, A7.

mean of x-coordinates

$$\frac{2+1}{2} = 1.5.$$

mean of y-coordinates.

$$\frac{5+2}{2} = 3.5.$$

$C_3(1.5, 3.5)$.

Distance calculation:

point	coordinates	Distance to c_1 (3, 9.5)	Distance to c_2 (6.5, 5.25)	Distance to c_3 (1.5, 3.5)	Nearest clusters
A1	(2,10)	1.12	6.02	6.80	c1
A2	(2,5)	4.53	4.61	1.58	c3
A3	(8,4)	7.43	1.68	6.60	c2
A4	(5,8)	2.69	2.96	5.41	c1
A5	(7,5)	5.41	0.79	5.70	c2
A6	(6,4)	5.59	1.27	4.61	c2
A7	(1,2)	7.59	6.10	1.58	c3
A8	(4,9)	1.127	3.91	5.70	c1

update cluster centers:

=) C1 : A1, A4, A8.

mean x-coordinates:

$$\frac{2+5+4}{3} = 3.67.$$

mean of y-coordinates:

$$\frac{10+8+9}{3} = 9.$$

c1 (3.67, 9).

=) C2 : A3, A5, A6.

mean of x-coordinates:

$$\frac{8+7+6}{3} = 7.$$

mean of y-coordinates:

$$\frac{4+5+4}{3} = 4.33.$$

$C_2(7, 4.33)$.

⇒ C_3 : A_2, A_7.

mean of x-coordinates:

$$\frac{2+1}{2} = 1.5.$$

mean of y-coordinates.

$$\frac{5+2}{2} = 3.5.$$

C_3 (1.5, 3.5)

After calculating the distance for each point we get the same result as previous one, so we stop.

Final clusters:

⇒ C1 : A1 (2, 10), A4 (5, 8), A8 (4, 9).
⇒ C2 : A3 (8, 4), A5 (7, 5), A6 (6, 4).
⇒ C3 : A2 (2, 5), A7 (1, 2).

final cluster centers:

⇒ C1 : (3.67, 9).
⇒ C2 : (7, 4.33)
⇒ C3 : (1.5, 3.5)