K-means Clustering

Q

Individual	Variable 1	Variable 2
→ 1	1.0	1.0
2	1.5	2.0
3	3.0	4.0
→ 4	5.0	7.0
5	3.5	5.0
6	4.5	5.0
7	3.5	4.5

K = 2

Sol:- Randomly we choose following 2 centroids (K=2) for 2 clusters.
In this case the 2 centroid are -

$$\boxed{\begin{aligned} m_1 &= (1.0, 1.0) \\ m_2 &= (5.0, 7.0) \end{aligned}}$$

Now calculating for individual 1

$$d(m_1, 1) = \sqrt{(1.0-1.0)^2 + (1.0-1.0)^2}$$
$$= \sqrt{0+0}$$
$$= 0$$

$$d(m_2, 1) = \sqrt{(1.0-5.0)^2 + (1.0-7.0)^2}$$
$$= \sqrt{4^2 + 6^2}$$
$$= \sqrt{16+36}$$
$$= \sqrt{52} = 7.21$$

Here the lowest distance comes with m_1 so, 1 goes with m_1

⇒ ~~Now, we need to update m_1 centroid~~

~~$m_1 = (1.0, 1.0)$ $m_2 = (5.0, 7.0)$~~

Page No. :
Date : / /

⇒ Now calculating for Individual 2

$d(m_1, 2) = \sqrt{(1.5-1.0)^2 + (2.0-1.0)^2}$
$= \sqrt{0.5^2 + 1.0^2}$
$= 1.12$

$d(m_2, 2) = \sqrt{(1.5-5.0)^2 + (2.0-7.0)^2}$
$= \sqrt{(-3.5)^2 + (-5)^2}$
$= 6.10$

Here lowest distance comes with m1
so, 2 goes to m1

~~Now we need to update the m_1 centroid~~

~~$m_1 = \frac{1.0+1.5}{2}, \frac{1.0+2.0}{2}$~~
~~$= \frac{2.5}{2}, \frac{3.0}{2}$~~
~~$m_1 = 1.25, 1.5$~~ ~~$m_2 = 5.0, 7.0$~~

⇒ Now calculating for individual 3 -

~~$d(m_1, 3) = \sqrt{(3.0-1.25)^2 + (4.0-1.5)^2}$~~
~~$= \sqrt{(1.75)^2 + (2.5)^2}$~~
~~$= \varnothing$~~

$d(m_1, 3) = \sqrt{(3.0-1.0)^2 + (4.0-1.0)^2}$
$= \sqrt{2^2 + 3^2}$
$= \sqrt{4+9} = \sqrt{13} = 3.61$

$d(m_2, 3) = \sqrt{(5.0-3.0)^2 + (7.0-4.0)^2}$
$= \sqrt{2^2 + 3^2}$
$= \sqrt{13} = 3.61$

Here distance of both are same, so

(3) we go with cluster m1

Now we need to update m1 centroid

$m_1 = 1.25$

⇒ Now calculating for individual 4-

$d(m_1, 4) = \sqrt{(5.0-1.0)^2 + (7.0-1.0)^2}$
$= \sqrt{4^2 + 6^2}$
$= \sqrt{16+36} = \sqrt{52}$
$= 7.21$

$d(m_2, 4) = \sqrt{(5.0-5.0)^2 + (7.0-7.0)^2}$
$= \sqrt{0+0} = 0$

Here lowest distance is with m_2 so,

(4) goes with m_2

⇒ Now calculating for individual 5-

$d(m_1, 5) = \sqrt{(3.5-1.0)^2 + (5.0-1.0)^2}$
$= \sqrt{2.5^2 + 4^2}$
$= 4.72$

$d(m_2, 5) = \sqrt{(3.5-5.0)^2 + (5.0-7.0)^2}$
$= \sqrt{(1.5)^2 + (-2)^2}$
$= \cancel{\sqrt{}}\ 2.5$

Here lowest distance is with m_2 So,

(5) goes with m_2

Now calculating for individual 6

$d(m_1, 6) = \sqrt{(4.5 - 1.0)^2 + (5.0 - 1.0)^2}$
$= \sqrt{3.5^2 + 4^2}$
$= 5.31$

$d(m_2, 6) = \sqrt{(4.5 - 5.0)^2 + (5.0 - 7.0)^2}$
$= \sqrt{-0.5^2 + (-2)^2}$
$= 2.06$

Here lower distance is with m_2, so, 6 goes with m_2

Now calculating for individual 7

$d(m_1, 7) = \sqrt{(3.5 - 1.0)^2 + (4.5 - 1.0)^2}$
$= \sqrt{2.5^2 + 3.5^2}$
$= 4.30$

$d(m_2, 7) = \sqrt{(3.5 - 5.0)^2 + (4.5 - 7.0)^2}$
$= \sqrt{-1.5^2 + (-2.5)^2}$
$= 2.92$

Here lower distance is with m_2, so 6 goes with m2

$m_1 = 1, 2, 3$
$m_2 = 4, 5, 6, 7$

Page No.
Date: / /

m_1 centroid :-

$$\frac{1.0+1.5+3.0}{3}, \frac{1.0+2.0+4.0}{3}$$

$\Rightarrow$ $\boxed{m_1 = 1.83, 2.33}$

m_2 Centroid :-

$$\frac{5.0+3.5+4.5+3.5}{4}, \frac{7.0+5.0+5.0+4.5}{4}$$

$\Rightarrow$ $\boxed{m_2 = 4.12, 5.38}$

Step 3 Now using these centroids we compute the Euclidean distance of each object.

Calculating for individual 1

$$d(m_1, 1) = \sqrt{(1.83-1.0)^2 + (2.33-1.0)^2}$$
$$= \sqrt{0.83^2 + (1.33)^2}$$
$$= \sqrt{} = 1.57$$

$$d(m_2, 1) = \sqrt{(4.12-1.0)^2 + (5.38-1.0)^2}$$
$$= \sqrt{3.12^2 + 4.38^2}$$
$$= 5.38$$

$\Rightarrow$ It goes with m_1

Calculating for Individual 2

$$d(m_1, 2) = \sqrt{(1.83-1.5)^2 + (2.33-2.0)^2}$$
$$= \sqrt{(0.33)^2 + (0.33)^2}$$
$$= 0.47$$

$$d(m_2, 2) = \sqrt{(4.12-1.5)^2 + (5.38-2.0)^2}$$
$$= \sqrt{2.62^2 + 3.38^2}$$
$$= 4.28$$

Page No.:
Date: / /

Here lower distance is with m1, so, (2) goes with m1

⇒ Calculating for individual 3

$$d(m_1,3) = \sqrt{(1.83-3.0)^2 + (2.33-4.0)^2}$$
$$= \sqrt{-1.17^2 + (-1.67)^2}$$
$$= 2.04$$

$$d(m_2,3) = \sqrt{(4.12-3.0)^2 + (5.38-4.0)^2}$$
$$= \sqrt{1.12^2 + 1.38^2}$$
$$= 1.78$$

Here lower distance is with m_2 so, (3) goes with m_2

⇒ Calculating for individual 4

$$d(m_1,4) = \sqrt{(1.83-5.0)^2 + (2.33-7.0)^2}$$
$$= \sqrt{-3.17^2 + (-4.67)^2}$$
$$= 5.64$$

$$d(m_2,4) = \sqrt{(4.12-5.0)^2 + (5.38-7.0)^2}$$
$$= \sqrt{(-0.88)^2 + (-1.62)^2}$$
$$= 1.84$$

Here lower distance is with m_2, so, (4) goes with m2

Page No.:
Date: / /

⇒ Calculating for individual 5

$d(m_1, 5) = \sqrt{(1.83-3.5)^2 + (2.33-5.0)^2}$
$= \sqrt{(-1.67)^2 + (-2.67)^2}$
$= 3.15$

$d(m_2, 5) = \sqrt{(4.12-3.5)^2 + (5.38-5.0)^2}$
$= \sqrt{(0.62)^2 + (0.38)^2}$
$= 0.73$

Here lower distance is with m_2, so, 5 goes with m_2

⇒ Calculating for individual 6

$d(m_1, 6) = \sqrt{(1.83-4.5)^2 + (2.33-5.0)^2}$
$= \sqrt{(-2.67)^2 + (-2.67)^2}$
$= 3.78$

$d(m_2, 6) = \sqrt{(4.12-4.5)^2 + (5.38-5.0)^2}$
$= \sqrt{(-0.38)^2 + (0.38)^2}$
$= 0.54$

Here lower distance is with m_2, so 6 goes with m_2

⇒ Calculating for individual 7

$d(m_1, 7) = \sqrt{(1.83-3.5)^2 + (2.33-4.5)^2}$
$= \sqrt{(-1.67)^2 + (-2.17)^2}$
$= 2.74$

$d(m_2, 7) = \sqrt{(4.12-3.5)^2 + (5.38-4.5)^2}$
$= \sqrt{0.62^2 + 0.88^2}$
$= 1.08$

Here, lower distance is with m_2, so 7 goes with m_2

New Clusters are = {1, 2} — m_1
{3, 4, 5, 6, 7} — m_2

Next Centroid are:-

$$m_1 = \frac{1.0+1.5}{2}, \frac{1.0+2.0}{2}$$

$$m_1 = 1.25, 1.5$$

$$m_2 = \frac{3.0+5.0+3.5+4.5+3.5}{5}, \frac{4.0+7.0+5.0+5.0+4}{5}$$

$$m_2 = 3.9, 5.1$$

Step 4

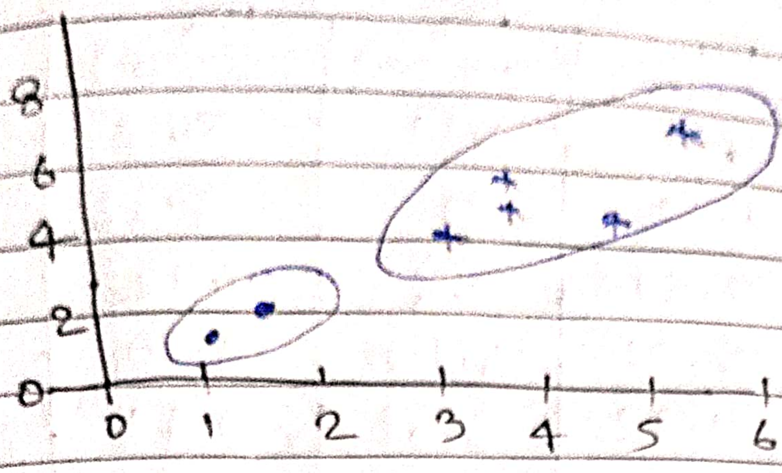
Now, using this centroid we again calculate the euclidean distance of each object

individual	Centroid 1	Centroid 2
1	0.58	0.52
2	0.58	3.92
3	3.05	1.42
4	6.66	2.20
5	4.16	0.41
6	4.78	0.61
7	3.75	0.72

The clustered obtained are:- {1, 2} — m_1
{3, 4, 5, 6, 7} — m_2

Therefore, there is no change in the cluster.
Thus, the algo. comes to a halt here and final result consist of 2 clusters {1, 2} and {3, 4, 5, 6, 7}

Plot



$\Rightarrow$ with $\underline{\underline{k=3}}$

Step-1 we randomly choose 3 centroids -

$m_1 = (1.0, 1.0)$ —— ①

$m_2 = (1.5, 2.0)$ —— ②

$m_3 = (3.9, 5.1)$ —— 3, 4, 5, 6, 7.

$$\frac{3.0+5.0+3.5+4.5+3.5}{5}, \frac{4.0+7.0+5.0+5.0+4.5}{5}$$

$m_3 = [\ 3.9\ ,\ 5.1\ ]$

Individual	$m_1 = 1$	$m_2 = 2$	$m_3 = 3$	cluster
1	0	1.11	3.61	1
2	1.12	0	2.5	2
3	3.61	2.5	0	3
4	7.21	6.10	3.61	3
5	4.72	3.61	1.12	3
6	5.31	4.24	1.80	3
7	4.30	3.20	0.71	3

C_3

Clustering with initial centroid (1, 2, 3)

Page No. :
Date : / /

Step - 2

Individual	m_1 (1.0, 1.0)	m_2 (1.5, 2.0)	m_3 (3.9, 5.1)	cluster
1	0	1.11	5.02	1
2	1.12	0	3.92	2
3	3.61	2.5	1.42	3
4	7.21	6.10	2.20	3
5	4.72	3.61	0.41	3
6	5.31	4.24	0.61	3
7	4.30	3.20	0.72	3

Plot