K Means Clustering

* Suppose that the ~~data~~ task is to cluster points into 3 clusters,

A1(2,10) A2(2,5) A3(8,4)
B1(5,8) B2(7,5) B3(6,4)
C1(1,2), C2(4,9)

* The distance function is Euclidean distance
* Suppose initially we assign A_1, B_1, C_1 as the center of each cluster respectively.
* If initial centeroids are not given, we can choose any points as initial centeroids.

Initial Centeroids: A1(2,10) B1(5,8) C1(1,2)

Data Points	x_1	y_1	Distance to (x_2 2, y_2 10)	Distance to (5, 8)	Distance to (1, 2)	cluster	New cluster
A_1	2	10	0	3.60	8.06	1	
A_2	2	5	5	4.24	3.16	3	
A_3	8	4	8.49	5	7.28	2	
B_1	5	8	3.60	0	7.21	2	
B_2	7	5	7.07	3.60	6.71	2	
B_3	6	4	7.21	4.12	5.39	2	
C_1	1	2	8.06	7.21	0	3	
C_2	4	9	2.24	1.41	7.62	2	

Euclidean distance formula

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

A1

A1 → $\sqrt{(2-2)^2 + (10-10)^2} = 0$

A2 → $\sqrt{(2-2)^2 + (10-5)^2} = \sqrt{25} = 5$

A3 → $\sqrt{(2-8)^2 + (10-4)^2} = \sqrt{36+36} = 8.49$

B1 → $\sqrt{(2-5)^2 + (10-8)^2} = \sqrt{9+4} = 3.60$

B2 → $\sqrt{(2-7)^2 + (10-5)^2} = \sqrt{25+25} = 7.07$

B3 → $\sqrt{(2-6)^2 + (10-4)^2} = \sqrt{16+36} = 7.21$

C1 → $\sqrt{(2-1)^2+(10-2)^2} = \sqrt{1+64} = 8.06$

C2 → $\sqrt{(2-4)^2+(10-9)^2} = \sqrt{4+1} = 2.24$

B1

A1 → $\sqrt{(5-2)^2+(8-10)^2} = \sqrt{9+4} = 3.60$

A2 → $\sqrt{(5-2)^2+(8-5)^2} = \sqrt{9+9} = 4.24$

A3 → $\sqrt{(5-8)^2+(8-4)^2} = \sqrt{9+16} = 5$

B1 → $\sqrt{(5-5)^2+(8-8)^2} = 0$

B2 = $\sqrt{(5-7)^2+(8-5)^2} = \sqrt{4+9} = 3.60$

B3 = $\sqrt{(5-6)^2+(8-4)^2} = \sqrt{1+16} = 4.12$

C1 = $\sqrt{(5-1)^2+(8-2)^2} = \sqrt{16+36} = 7.21$

C2 = $\sqrt{(5-4)^2+(8-9)^2} = \sqrt{1+1} = 1.41$

C1

A1 → $\sqrt{(1-2)^2+(2-10)^2} = \sqrt{1+64} = 8.06$

A2 → $\sqrt{(1-2)^2+(2-5)^2} = \sqrt{1+9} = 3.16$

A3 → $\sqrt{(1-8)^2+(2-4)^2} = \sqrt{49+4} = 7.28$

B1 → $\sqrt{(1-5)^2+(2-8)^2} = \sqrt{16+36} = 7.21$

B2 → $\sqrt{(1-7)^2+(2-5)^2} = \sqrt{36+9} = 6.71$

B3 → $\sqrt{(1-6)^2+(2-4)^2} = \sqrt{25+4} = 5.39$

C1 → $\sqrt{(1-1)^2+(2-2)^2} = 0$

C2 → $\sqrt{(1-4)^2+(2-9)^2} = \sqrt{9+49} = 7.62$

Now We need to calculate new centeroids

cluster 1 consists of only 1 point ie (2,10)

New centeroids: so A1 c_1 = (2,10)

cluster 2 consists of points like

(8,4), (5,8), (7,5), (6,4), (4,9)

find the centeroids

$$\frac{8+5+7+6+4}{5} = \frac{30}{5} = 6$$

$$\frac{4+8+5+4+9}{5} = \frac{30}{5} = 6$$

(6,6)

cluster 3 consists of points

(2,5) and (1,2)

$$\frac{2+1}{2} = 1.5 \qquad \frac{5+2}{2} = 3.5$$

(1.5, 3.5)

New centeroids

A1: (2,10) B1 (6,6) C_1: (1.5, 3.5)

Data Points			Distance to (2,10)	Distance to (6,6)	Distance to (1.5,3.5)	from old table cluster	New cluster
A1	2	10	0	5.66	6.52	1	1
A2	2	5	5	4.12	1.58	3	3
A3	8	4	8.49	2.83	6.52	2	2
B1	5	8	3.60	2.24	5.70	2	2
B2	7	5	7.07	1.41	5.70	2	2
B3	6	4	7.21	2	4.53	2	2
C1	1	2	8.06	6.40	1.58	3	3
C2	4	9	2.24	3.60	6.04	(2) →	(1)

A1

A1 : $\sqrt{(2-2)^2+(10-10)^2} = 0$

A2 : $\sqrt{(2-2)^2+(10-5)^2} = \sqrt{25} = 5$

A3 : $\sqrt{(2-8)^2+(10-4)^2} = \sqrt{36+36} = \sqrt{72} = 8.49$

B1 : $\sqrt{(2-5)^2+(10-8)^2} = \sqrt{9+4} = \sqrt{13} = 3.60$

B2 : $\sqrt{(2-7)^2+(10-5)^2} = \sqrt{25+25} = \sqrt{50} = 7.07$

B3 : $\sqrt{(2-6)^2+(10-4)^2} = \sqrt{16+36} = \sqrt{52} = 7.21$

C1 : $\sqrt{(2-1)^2+(10-2)^2} = \sqrt{1+64} = \sqrt{65} = 8.06$

C2 : $\sqrt{(2-4)^2+(10-9)^2} = \sqrt{4+1^2} = \sqrt{5} = 2.24$

A1

A1 : $\sqrt{(6-2)^2+(6-10)^2} = \sqrt{16+16} = \sqrt{32} = 5.66$

A2 : $\sqrt{(6-2)^2+(6-5)^2} = \sqrt{16+1^2} = \sqrt{17} = 4.12$

A3 : $\sqrt{(6-8)^2+(6-4)^2} = \sqrt{4+4} = \sqrt{8} = 2.83$

B1 : $\sqrt{(6-5)^2+(6-8)^2} = \sqrt{1+4} = \sqrt{5} = 2.24$

B2 : $\sqrt{(6-7)^2+(6-5)^2} = \sqrt{1+1} = \sqrt{2} = 1.41$

B3 : $\sqrt{(6-6)^2+(6-4)^2} = \sqrt{4} = 2$

C1 : $\sqrt{(6-1)^2+(6-2)^2} = \sqrt{25+16} = \sqrt{41} = 6.40$

C2 = $\sqrt{(6-4)^2+(6-9)^2} = \sqrt{4+9} = \sqrt{13} = 3.60$

C1

A1 : $\sqrt{(1.5-2)^2+(3.5-10)^2} = \sqrt{(.5)^2+(6.5)^2} = \sqrt{0.25+42.25}$

$= \sqrt{42.5} = 6.52$

A2 : $\sqrt{(1.5-2)^2+(3.5-5)^2} = \sqrt{(0.5)^2+(1.5)^2} = 1.58$

A3 : $\sqrt{(1.5-8)^2+(3.5-4)^2} = \sqrt{(-6.5)^2+(.5)^2} = 6.52$

B1 : $\sqrt{(1.5-5)^2+(3.5-8)^2} = \sqrt{12.25+20.25} = 5.70$

B2 : $\sqrt{(1.5-7)^2+(3.5-5)^2} = \sqrt{30.25+2.25} = 5.70$

B3 : $\sqrt{(1.5-6)^2+(3.5-4)^2} = \sqrt{20.25+0.25} = 4.53$

C1 : $\sqrt{(1.5-1)^2+(3.5-2)^2} = \sqrt{.25+2.25} = 1.58$

C2 : $\sqrt{(1.5-4)^2+(3.5-9)^2} = \sqrt{6.25+30.25} = 6.04$

C2 was assigned to 2nd cluster in previous case that Now it is moved to 1st cluster.

so this is not correct
we need to calculate new centeroids

Cluster 1 consists of 2 points
(2,10) (4,9)

$\frac{2+4}{2}$, $\frac{10+9}{2}$

$= 3$ 9.5

Cluster 2 consists of 4 points
(8,4) (5,8) (7,5) , (6,4)

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

$= \frac{26}{4}$ $= \frac{21}{4}$

$= 6.5$ $= 5.25$

Cluster 3 consists of 2 points.
(2,5) (1,2)

$\frac{2+1}{2}$ $\frac{5+2}{2}$

$= 1.5$ 3.5

Current Centeroids

A1 : (3, 9.5)
B1 : (6.5, 5.25)
C1 : (1.5, 3.5)

Data Points			Distance to 3 9.5	Distance to 6.5 5.25	Distance to 1.5 3.5	cluster	New cluster
A1	2	10	1.12	6.54	6.52	1	1
A2	2	5	4.61	4.51	1.58	3	3
A3	8	4	7.43	1.95	6.52	2	2
B1	5	8	2.50	3.13	5.70	2 →	1
B2	7	5	6.02	0.56	5.70	2	2
B3	6	4	6.26	1.35	4.53	2	2
C1	1	2	7.76	6.39	1.58	3	3
C2	4	9	1.12	4.51	6.04	1	1

New Centeroids

A1: 3.67, 9

B1: 7, 4.33

C1: 1.5, 3.5

Data Points			Distance to 3.67 9	Distance to 7 4.33	Distance to 1.5 3.5	cluster	new cluster
A1	2	10	1.94	7.56	6.52	1	1
A2	2	5	4.33	5.04	1.58	3	3
A3	8	4	6.62	1.05	6.52	2	2
B1	5	8	1.67	4.18	5.70	1	1
B2	7	5	5.21	0.67	5.70	2	2
B3	6	4	5.52	1.05	4.53	2	2
C1	1	2	7.49	6.44	1.58	3	3
C2	4	9	0.33	5.55	6.04	1	1

After cluster assignment, previous & new clusters are same ⇒ all datapts are converged to new clusters

A1 – 1 A2 — 3 A3, B2, B3 } – C2

B1 – 1 C1 — 3

C2 — 1