CS-559 Machine Learning

Homework Assignment-4

Problem 1: K-Means

⇒ The distance function of Euclidean distance is given by :-

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

⇒ Here, A_1, B_1 and C_2 is initially assigned the center of each cluster. (centroid)

Data Points			Distance to (A1) 2, 10	(B1) 5, 8	(C2) 4, 9	Cluster
A1	2	10	0	3.61	2.23	1
A2	2	5	5	4.24	4.47	2
A3	8	4	8.49	5	6.40	2
B1	5	8	3.61	0	1.41	2
C1	1	2	8.06	7.21	7.61	2
C2	4	9	2.23	1.41	0	3

Calculating the euclidean distance of each point to the initially assigned centroids.

$$d(A_1,A_1) = \sqrt{(2-2)^2+(10-10)^2} = 0$$

$$d(A_2,A_1) = \sqrt{(2-2)^2+(5-10)^2} = 5$$

$$d(A_3,A_1) = \sqrt{(8-2)^2+(4-10)^2} = 8.49$$

$$d(B_1, A_1) = \sqrt{(5-2)^2+(8-10)^2} = 3.61$$

$$d(C_1, A_1) = \sqrt{(1-2)^2+(2-10)^2} = 8.06$$

$$d(C_2, A_1) = \sqrt{(4-2)^2+(9-10)^2} = 2.23$$

$$d(A_1, B_1) = \sqrt{(2-5)^2+(10-8)^2} = 3.61$$

$$d(A_2, B_1) = \sqrt{(2-5)^2+(5-8)^2} = 4.24$$

$$d(A_3, B_1) = \sqrt{(8-5)^2+(4-8)^2} = 5$$

$$d(\cancel{A}B_1, B_1) = \sqrt{(5-5)^2+(8-8)^2} = 0$$

$$d(C_1, B_1) = \sqrt{(1-5)^2+(2-8)^2} = 7.21$$

$$d(C_2, B_1) = \sqrt{(4-5)^2+(9-8)^2} = 1.414$$

$$d(A_1, C_2) = \sqrt{(2-4)^2+(10-9)^2} = 2.23$$

$$d(A_2, C_2) = \sqrt{(2-4)^2+(5-9)^2} = 4.47$$

$$d(A_3, C_2) = \sqrt{(8-4)^2+(4-9)^2} = 6.40$$

$$d(B_1, C_2) = \sqrt{(5-4)^2+(8-9)^2} = 1.41$$

$$d(C_1, C_2) = \sqrt{(1-4)^2+(2-9)^2} = 7.61$$

$$d(C_2, C_2) = \sqrt{(4-4)^2+(9-9)^2} = 0$$

---

1) The cluster assignment of each data point after the first iteration is -

⇒ $A1 = (2,10)$ belongs to cluster 1 with the centroid $A1$,

⇒ $A_2 = (2,5)$, $A_3 = (8,4)$, $B_1 = (5,8)$ and $C_1 = (1,2)$ belongs to cluster 2 with the centroid as B_1

⇒ $C_2 = (4,9)$ belongs to cluster 3 with the centroid C_2.

2) The centroid after the first iteration,

- $A1 = (2,10)$

- For B_1,

$$x = \frac{2+8+5+1}{4} = 4$$

$$y = \frac{5+4+8+2}{4} = 4.75$$

It is $B_1 = (4, 4.75)$

- $C_2 = (4,9)$