Name : Danijal Saeed
Sap id: 53937
Section: BS Data Scienxe
Course: AI

Assignment 4

Exp 1:

Data in given Table 1 Decide into 3 clusters (i-e K:3) The intial values of C1, C2, C3 are 20, 40 and 80 respectively

Solution: Iteration # 1

10 = 4 times	50 = 3 times	150 = 3 times
25 = 3 times	80 = 4 times	
20 = 4 times	100 = 4 times	250 = 3 times

10	20	25	100	50	50	50
10	20	25	80	150	150	150
10	20	25	80	100	100	100
10	20	80	80	250	250	50

given: $C_1 = 20$, $C_2 = 40$, $C_3 = 80$

Iteration # 1

freq	Data items (x_i)	K1: C1=20 $(C_1 - x_i)$	K2: C2=40 $Dis(C_2 - x_i)$	K3: C3 $Dis(C_3 - x_i)$	Clusters
4	10	10	30	70	C1
4	20	0	20	60	C1
3	25	5	15	55	C1
3	50	30	10	30	C2
4	80	60	40	0	C3
4	100	80	60	20	C3
3	150	130	110	70	C3
3	250	230	210	170	C3

New $C_1 = \frac{(10\times4)+(20\times4)+(25\times3)}{11} = 17+2$

& New $C_2 = \frac{50\times3}{3} = 50$

New $C_3 = \frac{(150\times4)+(100\times4)+(150\times3)+(250\times3)}{14} = 137.142$

Iteration # 2

freq	x_i	K1: C1=8 $Dis\ \lvert C_1 - x_i \rvert$	K2: C2=50 $Dis\ \lvert C_2 - x_i \rvert$	K3: C3: 37 $Dis(C_3 - x_i)$	Clusters
4	10	8	40	127	C1
4	20	2	30	117	C1
3	25	7	25	112	C1
3	60	32	0	87	C2
4	80	62	30	57	C2
4	100	82	50	37	C3
3	150	132	100	13	C3
3	250	232	200	113	C3

New $C_1 = \frac{(10\times4)+(20\times4)+(25\times3)}{11} = 17.72$

New $C_2 = \frac{(30 \times 3) + (80 \times 4)}{7} = 67.14$

New $C_3 = \frac{(100 \times 4) + (150 \times 3) + (250 \times 3)}{10} = 160$

Iteration 3

freq	x_i	K1 $C_1 = 8$ Dis $\lvert C_1 - x_i \rvert$	K2 $C_2 = 67$ Dis $\lvert C_2 - x_i \rvert$	K3 $C_3 = 160$ Dis $\lvert C_3 - x_i \rvert$	Clusters
4	10	8	57	150	C_1
4	20	2	47	140	C_1
3	25	7	42	135	C_1
3	50	32	17	110	C_2
4	80	62	13	80	C_2
4	100	82	33	60	C_3
3	150	132	83	10	C_3
3	250	232	183	90	C_3

New $C_1 = \frac{(10 \times 4) + (20 \times 4) + (20 \times 3)}{11} = 17.72$

New $C_2 = \frac{(50 \times 3) + (80 \times 4) + (100 \times 4)}{11} = 79.01$

New $C_3 = \frac{(150 \times 3) + (250 \times 3)}{6} = 200$

Iteration 4

freq	x_i	$k_1 \cdot c_1 = 180$ Dis $(c_1 - x_i)$	$k_2 \cdot c_2 = 79$ Dis $(c_2 - x_i)$	$k_3 = c_3 = 200$ Dis $(c_3 - x_i)$	Clusters
4	10	8	69	190	c_1
4	20	2	59	180	c_1
3	25	7	54	175	c_1
3	50	32	29	150	c_2
4	80	62	1	120	c_2
4	100	82	21	100	c_2
3	150	132	71	50	c_3
3	250	232	171	50	c_3

New $c_1 = \frac{(10 \times 4) + (20 \times 4) + (25 \times 3)}{11} = 17.72$

New $c_2 = \frac{(50 \times 3) + (80 \times 4) + (100 \times 4)}{11} = 79.09$

New $c_3 = \frac{(150 \times 3) + (250 \times 3)}{6} = 200$

Example : 2 Solution

Point	Dis mean 1	Dis mean 2	Dis mean 3	Cluster
A1 (2,10)	6	5	9	1
A2 (2,5)	5	6	4	3
A3 (5,8)	12	7	9	2
A4 (7,5)	5	0	10	2
A5 (6,4)	10	5	9	2
A6 (1,2)	10	5	7	2
A7 (8,4)	9	10	0	3
8 (4,9)	3	2	10	2

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

Step : 3 Recompute Means

Mean $(C_1) = (2, 10)$

Mean $(C_2) = \left(\frac{8+5+7+6+4}{5}, \frac{4+8+5+4+9}{5} \right) =$

Mean $(C_3) = \left(\frac{2+1}{2}, \frac{5+2}{2} = (1.5, 3.2) \right)$

New centroids:

$M_1 = (2, 10)$, $M_2 = (6, 6)$, M_3

Step 4 - Iteration 2 Assignment

Point	Dist to M_1	Dist to M_2	Dist to M_3	Clusters
A_1	0	8	7	1
A_2	5	5	2	3
A_3	12	4	7	2
A_4	5	3	8	2
A_5	10	2	7	2
A_6	10	2	5	2
A_7	9	4	2	3
A_8	3	5	3	1

Clusters after Iteration 2

$C_1 = \{A_1, A_8\}$
$C_2 = \{A_3, A_4, A_5, A_6\}$
$C_3 = \{A_2, A_7\}$

Check convergence