



K-MEANS

Subject: Data Warehousing & Mining

Sem: V

Topic:

MO JUNE 2021	Use K-Means algorithm to create 3 clusters for given set of values. [2, 3, 6, 8, 9, 12, 15, 18, 22]
	<u>Step 1</u> : Randomly choose 3 points as initial cluster center. $K=3$
	Let's choose 3 random values as initial mean values.
	$m_1 = 3$ $m_2 = 8$ $m_3 = 12$
	<u>Note</u> these initial clusters are chosen randomly.
	Now we will create distance matrix for every entry.
	We will calculate the distance of each point from the initial centres.
	The point with minimum distance to the respective mean, will belong to that cluster initially.

DISTANCE MATRIX $\rightarrow 1$

	Centroid 1=3	centroid 2=8	centroid 3=12	Alloted cluster
2	1	6	11	K ₁
3	0	5	9	K ₁
6	3	2	6	K ₂
8	5	0	4	K ₂
9	6	1	3	K ₂
12	9	4	0	K ₃
15	12	7	3	K ₃
18	15	10	6	K ₃
22	19	14	10	K ₃

So, now we have items in 3 clusters

$$K_1 = \{2, 3\}, \quad K_2 = \{6, 8, 9\}, \quad K_3 = \{12, 15, 18, 22\}$$

Iteration 1 ends

Iteration 2

Recalculate the centroid (ie cluster mean)

$$\text{centroid } K_1 = \frac{2+3}{2} = 2.5$$

$$\text{centroid } K_2 = \frac{6+8+9}{3} = 7.6$$

$$\text{centroid } K_3 = \frac{12+15+18+22}{4} = 16.75$$



Subject: Data Warehousing & Mining

Sem: V

Topic:

Again calculate the distance of each point from this newly generated mean.

DISTANCE MATRIX 2

		Centroid1 = 2.5	Centroid2 = 7.6	Centroid3 = 16.75
2	K ₁	0.5	5.6	14.75
3	K ₁	0.5	4.6	13.75
6	K ₂	4.5	1.6	10.75
8	K ₂	5.5	0.4	8.75
9	K ₂	6.5	1.4	7.75
12	K ₂	9.5	4.4	4.75
15	K ₃	12.5	7.4	1.75
18	K ₃	15.5	10.4	1.25
22	K ₃	19.5	14.4	5.25

following are the redefined clusters after iteration 2

$$K_1 = \{2, 3\} \quad K_2 = \{6, 8, 9, 12\} \quad K_3 = \{\cancel{15}, \cancel{18}, \cancel{22}\}$$

End of iteration 2

$$\{15, 18, 22\}$$

Iteration 3

Recalculating new centroids for newly created clusters

$$\text{centroid } K_1 = \frac{2+3}{2} = 2.5$$

$\{2, 3\}$

$$\text{centroid } K_2 = \frac{6+8+9+12}{4} = 8.75$$

$$\text{centroid } K_3 = \frac{15+18+22}{3} = 18.33$$

Generating Distance Matrix 3

	centroid 1 = 2.5	centroid 2 = 8.75	centroid 3 = 18.33	
2	0.5	6.75	16.3	K_1
3	0.5	3.75	13.3	K_1
6	3.5	2.75	12.3	K_2
8	5.5	0.75	10.3	K_2
9	6.5	0.25	9.3	K_2
12	9.5	3.25	6.3	K_2
15	12.5	6.25	3.3	K_3
18	15.5	9.25	0.3	K_3
22	19.5	13.25	3.7	K_3

$$K_1 = \{2, 3\}, K_2 = \{6, 8, 9, 12\}, K_3 = \{15, 18, 22\}$$

so we see here, the clusters remain unchanged,
we can stop here.