ML

LECTURE-26

BY
Dr. Ramesh Kumar Thakur
Assistant Professor (II)
School Of Computer Engineering

Previous Year Question (K - Means Clustering)

- 1. Divide the given sample data set into two(2) clusters using K means Algorithm by using Euclidean distance. Use Cluster 1 (185, 72) and Cluster 2 (170, 56) as initial cluster centers.

Height	Weight
185	72
170	56
168	60
179	68
182	72
188	77

- Ans: Computing the distances from initial cluster centers

xi	Data Point	Distance from cluster-1 (185,72)	Distance from cluster-2 (170,56)	Min Dist	Assigned Cluster
X1	(185,72)	0	21.93	0	Cluster-1
X2	(170,56)	21.93	0	0	Cluster-2
X3	(168,60)	20.81	4.47	4.47	Cluster-2
X4	(179,68)	7.21	15	7.21	Cluster-1
X5	(182,72)	3	20	3	Cluster-1
X6	(188,77)	5.83	27.66	5.83	Cluster-1

Previous Year Question (K - Means Clustering)

- The cluster centers are recalculated as follows:-
- Cluster-1 new center = {(185+179+182+188)/4,(72+68+72+77)/4} = (183.5,72.25)
- Cluster-2 new center = {(170+168)/2,(56+60)/2} = (169,58)
- Computing the distances from new cluster centers

xi	Data Point	Distance from cluster-1 (183.5,72.25)	Distance from cluster-2 (169,58)	Min Dist	Assigned Cluster
X1	(185,72)	1.52	21.26	1.52	Cluster-1
X2	(170,56)	21.13	2.24	2.24	Cluster-2
X3	(168,60)	19.76	2.24	2.24	Cluster-2
X4	(179,68)	6.19	14.14	6.19	Cluster-1
X5	(182,72)	1.52	19.10	1.52	Cluster-1
X6	(188,77)	6.54	26.87	6.54	Cluster-1

- Since there is no reassignment of data points to different clusters so we will stop the algorithm.
- The Final clusters are as follows:-
- Cluster-1={X1,X4,X5,X6} represented by v1 = (183.5,72.25)
- Cluster-2={X2,X3} represented by v2 = (169,58)

Previous Year Question (K - Means Clustering)

- 2. b) Apply K-means clustering algorithm on given data for K=3. Use C1(2), C2(16), C3(38) as initial cluster centers. Data: 2, 4, 6, 3, 31, 12, 15, 16, 38, 35, 14, 21, 23, 25, 33. [5]
- Ans:

Previous Year Question (K - Means Clustering)

- 3. (b) Cluster the given dataset of e-commerce sales with the following features:

Product_ID: ID of the product sold
Price: Price of the product (in dollars)
Quantity: Quantity of the product sold
Category: Category of the product (A, B, or C)

We want to use K-means clustering to group the products into three clusters based on their price and quantity. We will use K=3 for the K-means algorithm.

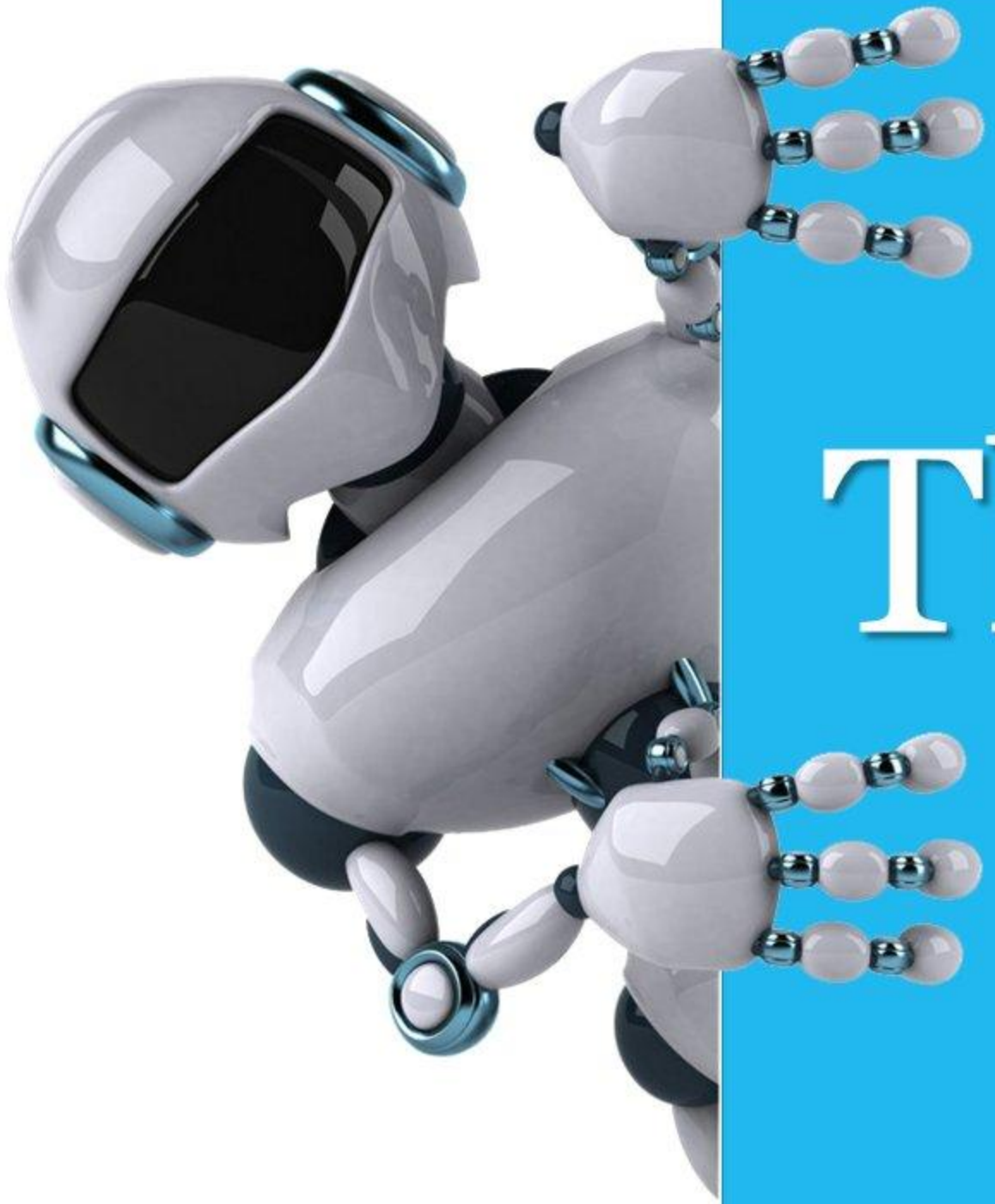
The sales data:

Product_ID	Price	Quantity	Category
1	50	10	A
2	40	8	A
3	60	12	A
4	70	15	B
5	80	18	B
6	90	20	C
7	100	22	C
8	110	25	C

- Ans:

Previous Year Question (K - Means Clustering)

- 4. 5. a) Apply K-means clustering algorithm on given data for K=2. Use C1(4), C2(12) as initial cluster centers. Data:{2, 3, 4, 10, 11, 12, 20, 25, 30} [3 Marks]
- Ans:



Thank you