SNS COLLEGE OF ENGINEERING

Kurumbapalayam(Po), Coimbatore – 641 107
Accredited by NAAC-UGC with 'A' Grade
Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai

Department of Information Technology

Course Name – Data Warehouse & Mining

II Year / IV Semester

Topic – Partitioning Method

Partitioning Method

S.No	Factor	K-Means	K-Medoids
1	Objective	Minimizing the sum of squared distances between data points and their assigned cluster centroids.	Minimizing the sum of dissimilarities between data points and their assigned cluster medoids.
2	Cluster Center Metric	Use centroids, which are the arithmetic means of all data points in a cluster.	Use medoids, which are representative data points within each cluster that are most centrally located concerning all other data points in the cluster.
3	Robustness	Less robust to noise and outliers.	More robust to noise and outliers.
4	Computational Complexity	Faster and more efficient for large datasets.	Slower and less efficient for large datasets.
5	Cluster Shape	Assumes spherical clusters and is not suitable for non-convex clusters.	Can handle non-convex clusters.
6	Initialization	Requires initial centroids to be randomly selected.	Requires initial medoids to be randomly selected.
7	Applications	Suitable for applications such as customer segmentation, image segmentation, and anomaly detection.	Suitable for applications where robustness to noise and outliers is important, such as clustering DNA sequences or gene expression data.

Example sum for Partitioning Method

- 2D data points

A(2, 10), B(2, 5), C(8, 4), D(5, 8), E(7, 5), F(6, 4)

Cluster these points into 2 clusters (k = 2) using the K-Means algorithm with the initial centroids:

Cluster 1 centroid = A(2, 10)

Cluster 2 centroid = C(8, 4)

Solution :

Step 1: Initial centroids

Cluster 1 (C1): (2, 10)

Cluster 2 (C2): (8, 4)

Example sum for Partitioning Method

Step 2: Assign each point to the nearest centroid (Euclidean distance)

Formula:
Distance=$\sqrt{(x2-x1)^2+(y2-y1)^2}$

Point	Distance to C1 (2,10)
A(2,10)	A(2,10)
B(2,5)	√25 = 5.0
C(8,4)	√(36+36)=√72 ≈ 8.49
D(5,8)	√(9+4)=√13 ≈ 3.61
E(7,5)	√(25+25)=√50 ≈ 7.07
F(6,4)	√(16+36)=√52 ≈ 7.21

Cluster assignment after Step 2:

Cluster 1: A, B, D

Cluster 2: C, E, F

Step 3: Recalculate centroids

New C1: Mean of A(2,10), B(2,5), D(5,8)

- $\bar{x}$ = (2+2+5)/3 = 3
- $\bar{y}$ = (10+5+8)/3 = 7.67

→ New C1 = (3, 7.67)

New C2: Mean of C(8,4), E(7,5), F(6,4)

- $\bar{x}$ = (8+7+6)/3 = 7
- $\bar{y}$ = (4+5+4)/3 ≈ 4.33

→ New C2 = (7, 4.33)

Categories of Clustering Method

Step 4: Reassign points (repeat distance check with new centroids)

You can continue until no changes in clusters (convergence), but usually 2–3 iterations are enough for small data.

Final Result:

After convergence, you'll have two distinct clusters:

- Cluster 1: A(2,10), B(2,5), D(5,8)
- Cluster 2: C(8,4), E(7,5), F(6,4)

THANK YOU