WHAT IS K-MEANS CLUSTERING?

- **Definition: K-Means is a clustering algorithm that groups data into kkk clusters based on similarity.**
-
- **Purpose: To partition the data such that points within each cluster are similar to each other and dissimilar to points in other clusters.**
-
-

~Applications
***Market segmentation**
***Image compression**
***Anomaly detection**

GAUSSIAN MIXTURE MODELS (GMM)

OVERVIEW:

- MODELS DATA AS A MIXTURE OF GAUSSIAN DISTRIBUTIONS.
- EACH CLUSTER IS REPRESENTED BY A GAUSSIAN COMPONENT.
- USES EXPECTATION-MAXIMIZATION (EM) ALGORITHM FOR PARAMETER ESTIMATION.

STRENGTHS: FLEXIBLE, CAN HANDLE OVERLAPPING CLUSTERS, PROBABILISTIC APPROACH

WEAKNESSES: CAN BE COMPUTATIONALLY EXPENSIVE, SENSITIVE TO INITIALIZATION

DBSCAN (DENSITY-BASED SPATIAL CLUSTERING OF APPLICATIONS WITH NOISE)

OVERVIEW:

- GROUPS TOGETHER POINTS THAT ARE CLOSELY PACKED TOGETHER.
- IDENTIFIES CLUSTERS BASED ON DENSITY THRESHOLDS CAN HANDLE CLUSTERS OF ARBITRARY SHAPE.

STRENGTHS: ROBUST TO NOISE, DISCOVERS CLUSTERS OF VARYING SHAPES AND SIZES

WEAKNESSES: SENSITIVE TO PARAMETER SELECTION (EPSILON AND MINPTS), CAN BE INEFFICIENT FOR LARGE DATASETS

COMPARISON: K-MEANS VS. GMM

Feature	K-Means	Gaussian Mixture Models
Model Assumption	Spherical clusters	Arbitrary-shaped clusters
Clustering Approach	Partitional	Probabilistic
Parameter Estimation	Simple	EM algorithm
Sensitivity to Initialization	Yes	Yes

COMPARISON: K-MEANS VS. DBSCAN

Feature	K-Means	DBSCAN
Cluster Shape	Spherical	Arbitrary
Noise Handling	Less robust	Robust
Parameter Sensitivity	Sensitive to K	Sensitive to epsilon and minPts
Outlier Detection	Implicit	Explicit

WHEN TO USE WHICH ALGORITHM:

K-MEANS: SIMPLE, EFFICIENT, WORKS WELL FOR WELL-SEPARATED CLUSTERS

GMM: FLEXIBLE, SUITABLE FOR OVERLAPPING CLUSTERS, PROBABILISTIC APPROACH

DBSCAN: ROBUST TO NOISE, HANDLES CLUSTERS OF ARBITRARY SHAPES, GOOD FOR OUTLIER DETECTION