Clustering Problem Definition

Core Objective

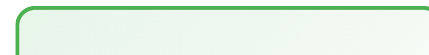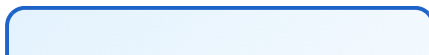🎯

Partition n observations into k groups

⚖️

Maximize intra-cluster similarity
Minimize inter-cluster similarity

Distance Metrics: Euclidean • Manhattan • Cosine • Mahalanobis

Types of Clustering Approaches

Hard Clustering

Each point belongs to exactly one cluster

Binary membership

Soft Clustering

Points have probability distribution over clusters

Fuzzy membership

Hierarchical

Nested clusters forming tree structure

Dendrogram output

Density-Based

Clusters as high-density regions

Arbitrary shapes

Model-Based Approach

Data generated from mixture of probability distributions