Unlocking Hidden Patterns

A Journey Through Clustering Algorithms

Discover how clustering algorithms find order in chaos, grouping similar data points together


Affinity Propagation


Mean Shift


Spectral


DBSCAN


OPTICS


BIRCH

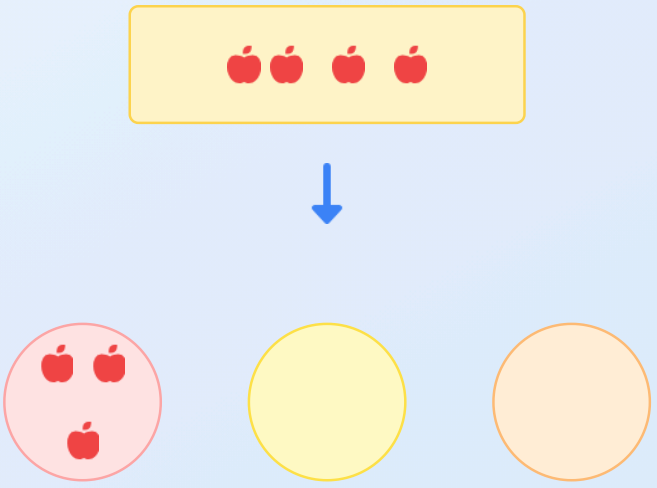
2025-11-25

Presented by: Harikrishnan

What is Clustering?

Clustering is like sorting mixed fruits into separate piles based on their similarities.

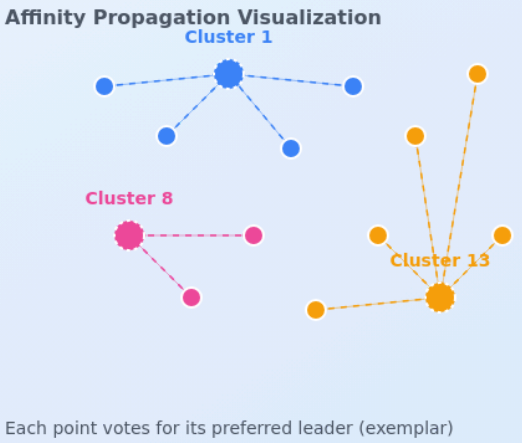
- It helps find hidden groups in data
- Groups similar items together
- Works without knowing groups beforehand



Affinity Propagation

Data points vote for exemplars like choosing group leaders at a party.

- Points exchange "votes" to find the best representatives
- Automatically determines optimal cluster numbers
- Each cluster has its own leader (exemplar)

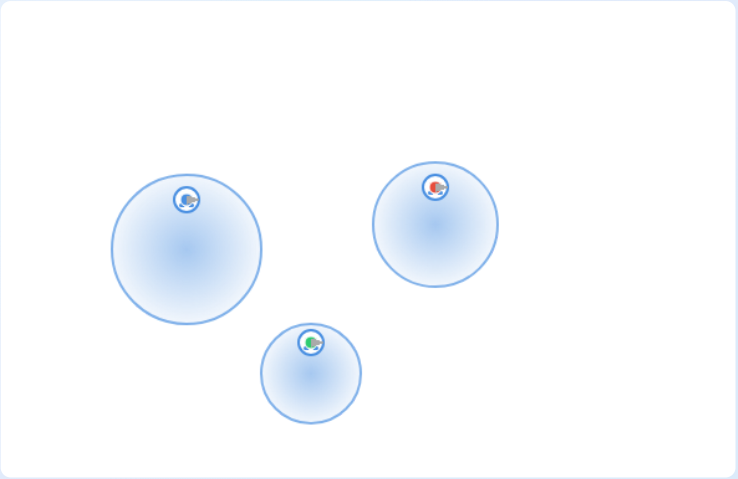


Each point votes for its preferred leader (exemplar)

Mean Shift

Points climb toward dense areas like reaching hilltops, grouping together those that reach the same peak.

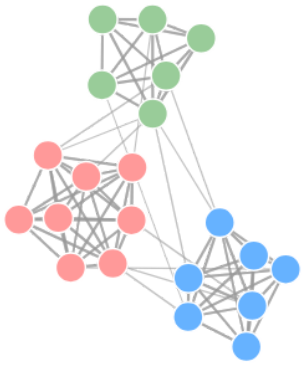
- Each point moves toward the average of its neighbors
- Points at the same peak form a cluster
- Finds clusters of different shapes



Spectral Clustering

Finds communities in networks by analyzing connections rather than just distance.

- Like finding friend groups in a social network
- Considers how well-connected points are
- Ideal for clusters with unusual shapes



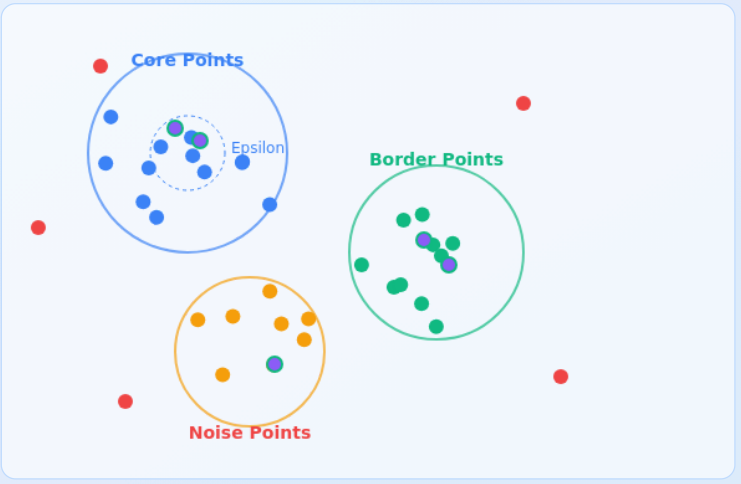
DBSCAN

"Crowded neighborhood" clustering

- Finds areas with many points nearby
- Points with enough neighbors form clusters
- Isolated points labeled as noise

Key Concepts:

- Core points
- Border points
- Noise points

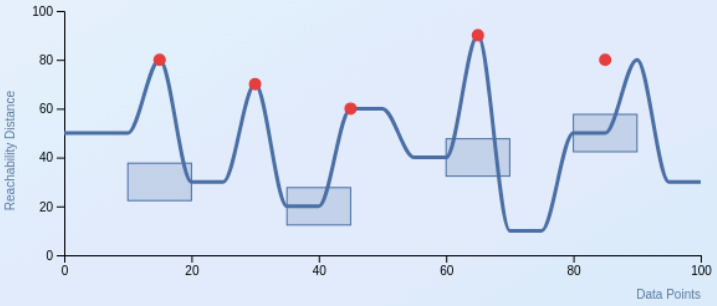


OPTICS

Advanced Neighborhood Analysis

OPTICS is like an "advanced neighborhood watch" that builds on DBSCAN, especially useful when some neighborhoods are much denser than others.

- Creates a "reachability map" showing how easily points connect
- Deep valleys represent dense clusters
- "Cuts" the map at different levels to find clusters



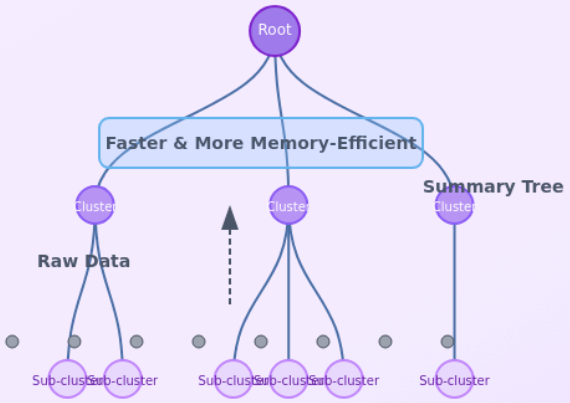
Hierarchical Structure

Various Density Levels

BIRCH

Efficiently handles massive datasets by creating summary trees, making clustering faster and more memory-efficient.

- Designed for massive datasets with millions of data points
- Creates a tree-like structure to summarize data
- Groups similar points into small sub-clusters
- Clusters summaries instead of individual points



Summary & Comparison

Algorithm	Core Idea	Best For
Affinity Propagation	Points "vote" for exemplars that receive most support	When you don't know how many clusters to expect
Mean Shift	Points move toward average of neighbors, clustering toward density peaks	Finding clusters of arbitrary shapes
Spectral Clustering	Uses graph structure to find communities with strong connections	Groups with unusual or intertwined shapes
DBSCAN	Finds "crowded neighborhoods" with core points and border points	Arbitrary shapes and noisy data
OPTICS	Creates "reachability map" showing density relationships	Clusters at various densities
BIRCH	Creates summary tree to efficiently cluster massive datasets	Very large datasets

Clustering helps find order in data by grouping similar items together.

💡 Choose algorithm based on data shape & goals!

2025-11-25

Presented by: Harikrishnan