DBSCAN - Density-Based Clustering

Density-Based Spatial Clustering of Applications with Noise

Parameters

ε (epsilon)
Neighborhood radius

minPts
Minimum density threshold

Core Points

≥ minPts neighbors within ε

Border Points

In neighborhood of core, but not core

Noise Points

Neither core nor border (outliers)

🎯 **Discovers clusters of arbitrary shape**

💪 Advantages

- ✓ Finds outliers automatically
- ✓ No need to specify k
- ✓ Handles arbitrary shapes
- ✓ Works with non-spherical clusters

⚠ Challenges

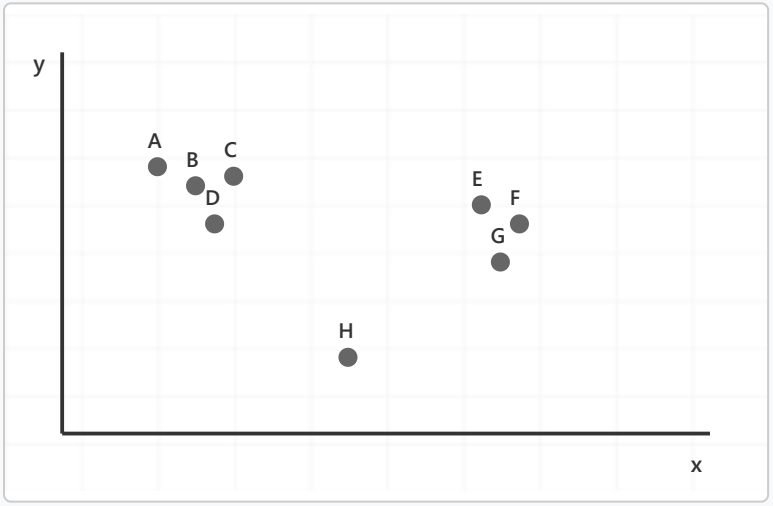
- ⚠ Sensitive to parameters (ε, minPts)
- ⚠ Struggles with varying densities
- ⚠ Difficulty in high dimensions

DBSCAN Step-by-Step Calculation Example

Detailed Process Using 2D Coordinate Data

Radius (Epsilon)	Minimum Points
ε = 1.5	**minPts = 3**

1 Dataset Definition



Point	x coord	y coord
A	2.0	3.0
B	2.5	2.5
C	3.0	2.7
D	2.8	1.8
E	6.5	2.5
F	7.0	2.0
G	6.8	1.5
H	5.0	0.5

Total 8 points

Data distributed on 2D plane

2 Distance Calculation - Euclidean Distance

Finding neighbors of Point A (2.0, 3.0)

$$d(A,B) = \sqrt{[(2.5-2.0)^2 + (2.5-3.0)^2]} = \sqrt{[0.25 + 0.25]} = 0.71 < 1.5 \checkmark$$

$$d(A,C) = \sqrt{[(3.0-2.0)^2 + (2.7-3.0)^2]} = \sqrt{[1.0 + 0.09]} = 1.04 < 1.5 \checkmark$$

$$d(A,D) = \sqrt{[(2.8-2.0)^2 + (1.8-3.0)^2]} = \sqrt{[0.64 + 1.44]} = 1.44 < 1.5 \checkmark$$

$$d(A,E) = \sqrt{[(6.5-2.0)^2 + (2.5-3.0)^2]} = \sqrt{[20.25 + 0.25]} = 4.53 > 1.5 \times$$

Neighbors of A: {B, C, D} → 3 points ≥ minPts(3) → Core Point

3

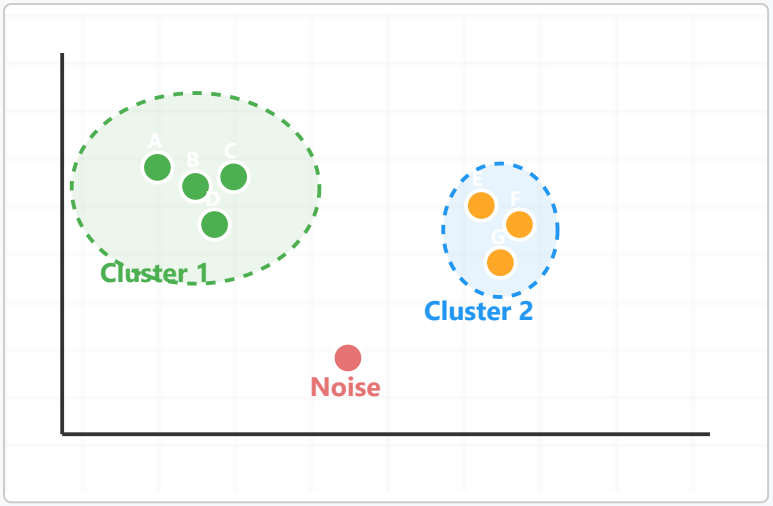
Finding Neighbors for All Points

Point	Neighbors (within ε=1.5)	# Neighbors	Type
A	{B, C, D}	3	Core
B	{A, C, D}	3	Core
C	{A, B, D}	3	Core
D	{A, B, C}	3	Core
E	{F, G}	2	Non-Core
F	{E, G}	2	Non-Core
G	{E, F}	2	Non-Core
H	{}	0	Noise

Core Points: A, B, C, D
Non-Core Points: E, F, G (neighbors < minPts)
Isolated Point: H (no neighbors)

4

Cluster Formation - Connectivity Check



Cluster 1 Formation:

- A is Core Point (starting point)
- Add neighbors B, C, D of A
- B, C, D are all Core Points
- Connected together to form one cluster

Cluster 2 Check:

- E, F, G are Non-Core Points
- Neighbors of each other but < minPts
- E, F, G are connected but
- No Core Point, needs special handling

H is not a neighbor of any Core Point → Noise

5

Final Classification Results

Point	Final Classification	Cluster	Description
A	Core Point	Cluster 1	Meets density condition, cluster center
B	Core Point	Cluster 1	Meets density condition, cluster center
C	Core Point	Cluster 1	Meets density condition, cluster center
D	Core Point	Cluster 1	Meets density condition, cluster center
E	Border Point	Cluster 2	Insufficient neighbors, connected to other Borders
F	Border Point	Cluster 2	Insufficient neighbors, connected to other Borders
G	Border Point	Cluster 2	Insufficient neighbors, connected to other Borders

Point	Final Classification	Cluster	Description

🎯 Clustering Results Summary

Cluster 1	Cluster 2	Noise
4 points	**3 points**	**1 point**
A, B, C, D (All Core)	E, F, G (All Border)	H (Outlier)