DBSCAN and HDBSCAN Clustering: Summary

# 1. DBSCAN (Density-Based Spatial Clustering of Applications with Noise)

DBSCAN is a density-based clustering algorithm that identifies clusters of arbitrary shape based on point density. It distinguishes between core points, border points, and noise.

## Key Concepts

- Core Point: Has at least 'n' points (including itself) within a radius ε.  
- Border Point: Within a core point's neighborhood but has fewer than 'n' neighbors.  
- Noise Point: Not a core or border point.

## Algorithm Steps

1. Choose minimum points (n) and radius (ε).  
2. For each point:  
 - If it has ≥ n points within ε → Core point.  
 - If it's in ε of a core point but has < n points → Border point.  
 - Otherwise → Noise.  
3. Expand clusters from core points, assigning border points.  
4. Unlabeled points are marked as noise.

# 2. HDBSCAN (Hierarchical DBSCAN)

HDBSCAN extends DBSCAN by removing the need for a fixed radius ε. It builds a hierarchy of density-based clusters and extracts the most stable ones.

## Key Features

- Does not require ε parameter.  
- More robust to noise and outliers.  
- Adapts to local density variations.  
- Uses cluster stability across density thresholds.

## Algorithm Steps

1. Treat each point as its own cluster (initially noise).  
2. Gradually merge clusters by lowering the density threshold.  
3. Build a hierarchy and extract stable clusters using a condensed tree.

# 3. DBSCAN vs. HDBSCAN

|  |  |  |
| --- | --- | --- |
| Feature | DBSCAN | HDBSCAN |
| Parameter Requirement | Requires ε and min\_samples | Only requires min\_cluster\_size |
| Outlier Handling | Good | Better |
| Performance on Varying Densities | Weak | Strong |
| Cluster Shape | Arbitrary | Arbitrary |
| Noise Tolerance | Moderate | High |

# 4. Use Cases

- Clustering real-world noisy data such as GPS coordinates.  
- Discovering arbitrarily shaped data clusters.  
- Pattern detection in geographical, biological, or customer data.  
- Applications where the number of clusters is unknown.