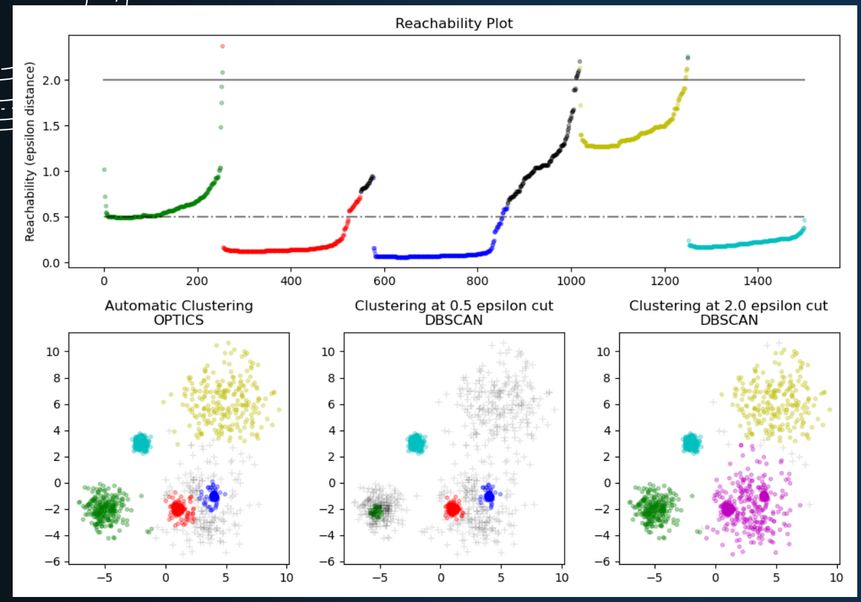
OPTICS Clustering

MACHINE LEARNING CLUSTERING

OPTICS Clustering

Introduction to OPTICS Clustering

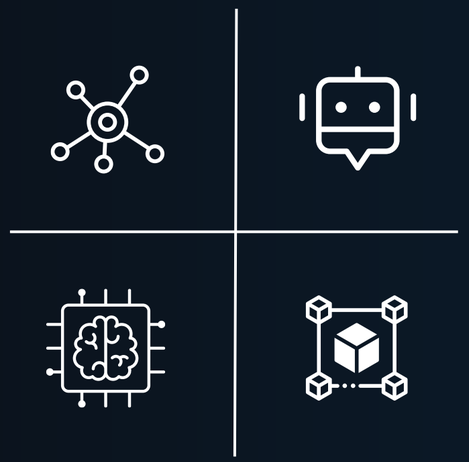
Ordering Points To Identify the Clustering Structure (OPTICS) is an algorithm that identifies clusters of varying density. Extends DBSCAN by creating an ordering of points based on density, which can be used to extract clusters at different density levels.



Key Concepts:

- **Epsilon (ε):** Maximum radius of the neighborhood around a point.
- **MinPts:** Minimum number of points required to form a dense region.
- **Core Distance:** Minimum distance ε' at which a point is a core point.
- **Reachability Distance:** Minimum distance a point needs to be reachable from a core point.

Mechanics of OPTICS Clustering



1. **Initialize:**
   - Start with an arbitrary point and mark it as visited.
2. **Core Distance Calculation:**
   - For each point, calculate the core distance if it has at least MinPts within ε.
3. **Expand Cluster Order:**
   - Use the core distance to determine the reachability distance of neighboring points.
   - Sort and add points to the ordering based on the reachability distance.
   - Expand the process to the next unvisited point with the smallest reachability distance.
4. **Extract Clusters:**
   - Use the reachability plot to identify clusters by identifying valleys in the plot, which represent dense regions.

Application and Evaluation

Application:

- Useful for data with varying densities, such as spatial data analysis and market segmentation.
- Steps:
  - ✓ Run OPTICS to generate an ordering of points and reachability distances.
  - ✓ Analyze the reachability plot to identify clusters.

Advantages:

- Identifies clusters of varying densities.
- Does not require the number of clusters to be specified.
- More robust to noise compared to DBSCAN.

Disadvantages:

- Computationally expensive, especially for large datasets.
- Sensitive to the choice of ε and MinPts.
- Interpretation of reachability plot can be complex.

Thank You