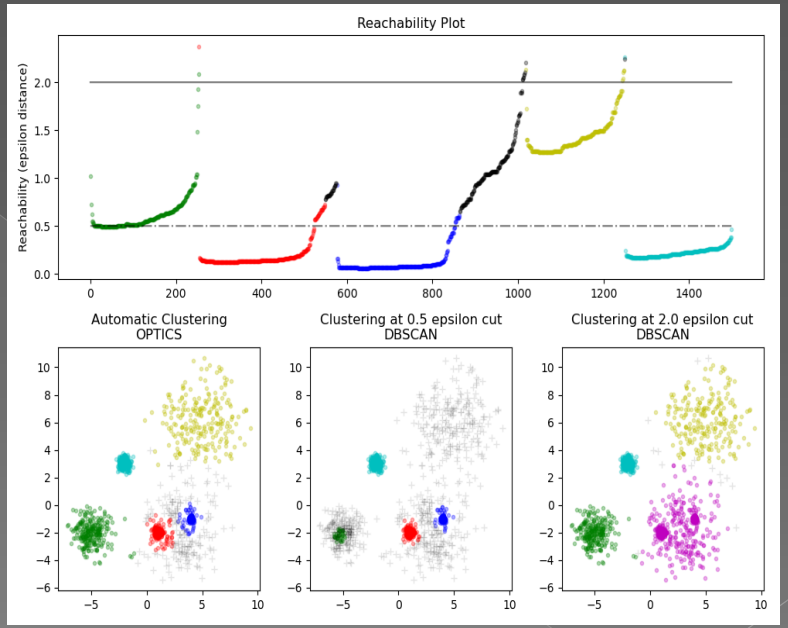
OPTICS Clustering

Ordering **P**oints **T**o **I**dentify the **C**lustering **S**tructure

- The OPTICS algorithm is a density-based clustering algorithm used for identifying clusters in spatial data.
- It extends upon the DBSCAN algorithm by providing a more flexible approach for discovering clusters of arbitrary shapes and sizes, as well as revealing the hierarchical structure of the data.



Key components

- Reachability Distance
- Core Distance
- Ordering of Points
- Clustering Structure

How Does OPTICS Work?

- Initialization
- Reachability Distance Calculation
- Point Ordering
- Cluster Extraction
- Hierarchical Clustering Structure
- Parameter Tuning
- Output

OPTICS Clustering v/s DBSCAN Clustering:

- Memory Cost
  - The OPTICS clustering technique requires more memory as it maintains a priority queue (Min Heap) is used to find Reachability Distance. Where as DBSCAN requires less memory space.
- Handling varying densities
  - OPTICS can identify clusters of different sizes and shapes more effectively than DBSCAN in datasets with varying densities.
- Noise handling
  - OPTICS may be less effective when compared to DBSCAN at identifying small clusters that are surrounded by noise points, as these clusters may be merged with the noise points in the reachability distance plot.
- Runtime complexity
  - The runtime complexity of OPTICS is generally higher than that of DBSCAN
- Fewer Parameters
  - OPTICS has fewer parameters when compared to DBSCAN

Advantage

- Ability to identify clusters of arbitrary shapes and sizes
- Robustness to noise and outliers
- Preservation of hierarchical clustering structure

Disadvantage

- Computational complexity, especially for large datasets
- Sensitivity to parameter selection
- Difficulty in handling high-dimensional data

Application

- Spatial data analysis
- Image segmentation
- Anomaly detection
- Customer segmentation