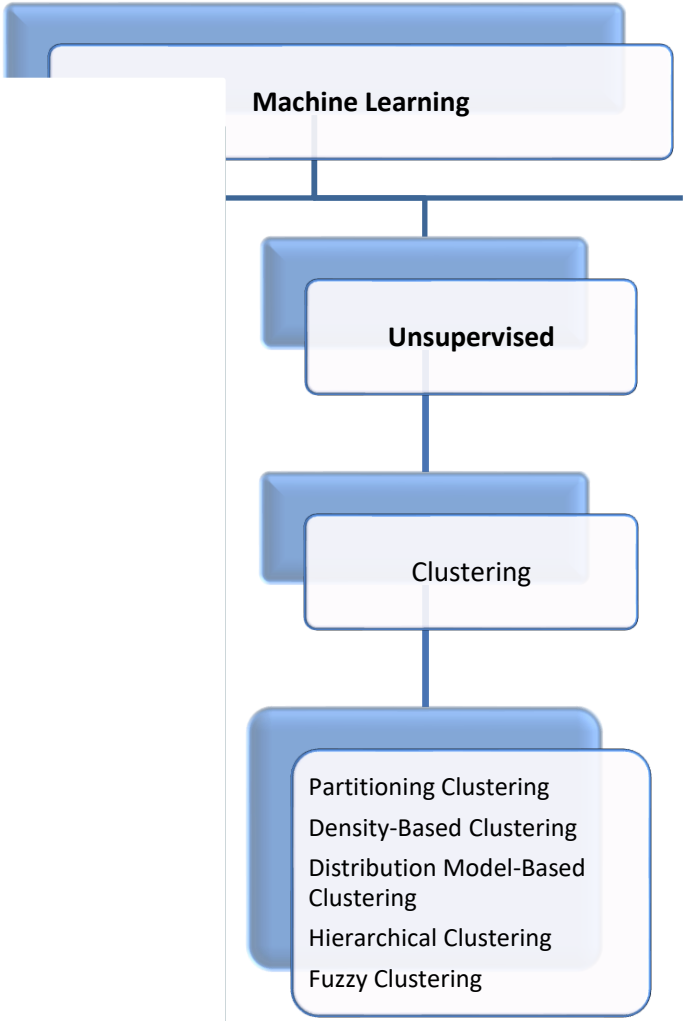
DBSCAN

MUKESH KUMAR


Machine Learning
Unsupervised
Clustering
Partitioning Clustering
Density-Based Clustering
Distribution Model-Based Clustering
Hierarchical Clustering
Fuzzy Clustering

Density-based methods

- DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a popular clustering algorithm used in data mining and machine learning for identifying clusters in a dataset. Unlike traditional clustering algorithms like K-means, which require the number of clusters to be specified beforehand, DBSCAN can automatically determine the number of clusters based on the data's density distribution

Advantages of DBSCAN

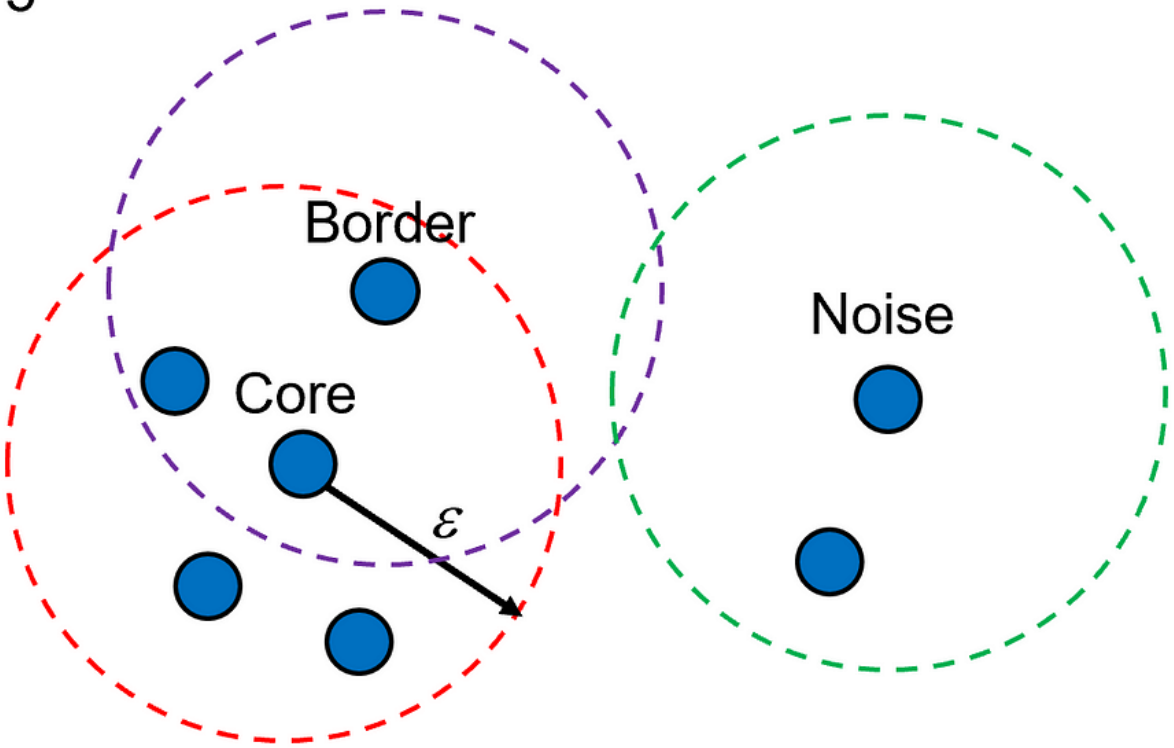
- **No Need for a Priori Specification of the Number of Clusters:** Unlike K-means, DBSCAN does not require specifying the number of clusters beforehand.
- **Ability to Find Arbitrarily Shaped Clusters:** DBSCAN can find clusters of various shapes and sizes, as it does not assume any specific cluster shape.
- **Robustness to Noise:** DBSCAN can identify and exclude noise points, making it robust to outliers.

Limitations of DBSCAN

- **Choice of Parameters:** The performance of DBSCAN heavily depends on the choice of ε and MinPts. Poor choices can lead to either too many clusters or merging of distinct clusters.

- **Varying Density:** DBSCAN struggles with datasets where clusters have varying densities, as a single ε value may not be appropriate for all clusters

MinPts = 5



Key Concepts

- **Epsilon (ε):** This is the maximum distance between two points for them to be considered as part of the same neighborhood.
- **MinPts (Minimum Points):** This is the minimum number of points required to form a dense region (or cluster).
- **Core Points:** Points that have at least MinPts points within a radius of ε.
- **Border Points:** Points that have fewer than MinPts within ε but are within the neighborhood of a core point.
- **Noise Points:** Points that are neither core points nor border points; these are considered outliers.

DBSCAN Algorithm

- Find all the neighbor points within eps and identify the core points than MinPts neighbors.
- For each core point if it is not already assigned to a cluster, create a new cluster.
- Find recursively all its density-connected points and assign them to the same cluster as the core point.
- Iterate through the remaining unvisited points in the dataset.
- Those points that do not belong to any cluster are noise.