DBSCAN CLUSTERING PYTHON

RAJAT SRIVASTAVA

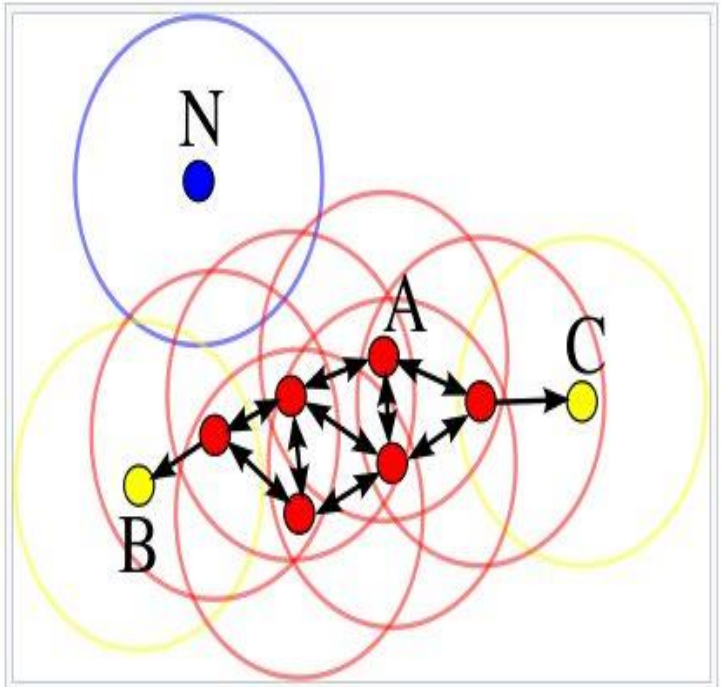
Clustering is applied on a dataset to group similar sets of data points. It looks for similarities and dissimilarities in data points and clutters them together. There are no labels in clustering. Clustering is an unsupervised learning to find the underlying structure of the dataset.

Types of clustering algorithms:

- Partition-based clustering
- Fuzzy clustering
- Hierarchical clustering
- Grid-Based clustering
- Exclusive clustering
- Overlapping clustering
- Density-based clustering

What is DBSCAN?

- DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a density-based unsupervised learning algorithm. It computes nearest neighbor graphs to find arbitrary-shaped clusters and outliers. Whereas the K-means clustering generates spherical-shaped clusters.
- DBSCAN does not require **K** clusters initially. Instead, it requires two parameters: eps and minPts.



- **eps**: it is the radius of specific neighborhoods. If the distance between two points is less than or equal to esp, it will be considered its neighbors.
- **minPts**: minimum number of data points in a given neighborhood to form the clusters.

- DBSCAN uses these two parameters to define a core point, border point, or outlier.

How does the DBSCAN clustering algorithm work?

- Randomly selecting any point **p**. It is also called **core point** if there are more data points than **minPts** in a neighborhood.
- It will use **eps** and **minPts** to identify all density reachable points.
- It will create a cluster using **eps** and **minPts** if **p** is a core point.
- It will move to the next data point if **p** is a **border point**. A data point is called a border point if it has fewer points than **minPts** in the neighborhood.
- The algorithm will continue until all points are visited.