**DBSCAN**

**Understanding DBSCAN:**

* **Concept**: DBSCAN is a density-based clustering algorithm that groups together points that are closely packed together, marking points in low-density regions as noise. It's effective for handling clusters of arbitrary shapes and can handle noise/outliers well.
* **Parameters**:
  + **eps**: The radius of the neighborhood around a point.
  + **min\_samples**: The minimum number of points required in a neighborhood for a point to be considered a core point (central to a cluster).
* **How It Works**: DBSCAN identifies core points (those with at least min\_samples points within eps distance), expands clusters from these points, and classifies points that don't belong to any cluster as noise.

**2. Advantages and Limitations:**

* **Advantages**:
  + Handles clusters of various shapes.
  + Robust to outliers.
* **Limitations**:
  + Struggles with varying densities in clusters.
  + Choosing optimal eps and min samples can be challenging.

Process explain : <https://www.youtube.com/watch?v=RDZUdRSDOok>

In simple terms :

We first decide a radius and make circles(region) and the no of points in the circle(Core point).

Then we find out the points that satisfy the condition of having minimum no of points(core points) in the region.

Then the points which satisfy these condition we mark it as a core point and other as a non core point.

Then we randomly select one core point and make it cluster and add the points to the cluster which are in that point region.

Then for the points which are added recently in the cluster we do the same process .

Note non-core points are not added now. After completing all the points in the cluster.

We add the non-core points which are in the region of these points in the cluster.

Note we can’t extend the points as we did for core points.

Then we go for other random point and do the same work.

This process continue until all the points are meted.

Note that point which are already considered in previous cluster will not be participate in new cluster.

The remaining points that are not in the comes in any cluster are marked as outliers.

Example :

Let Assume radius as 10 and core point as 4.

Then for each point in the graph we make a circle with radius of 10 and we figure out whether it has minimum of points in that region or not. If yes then it marked as core point else It is non-core point.

After figure out we pick random core point and make it as cluster. And add the points that are in that region ex : 7 points in that region. Then we add all the 7 points(only core points). Now in this cluster we have 8 points in out cluster.

And for the other 7 core points we do the same process. After completing all points. Then we add the non-core points that are in the region of these points.

Then we go and pick another point(in the remaining points) and do the same process.