

DBSCAN Clustering

Density Based Spatial Clustering of Applications with Noise

outliers

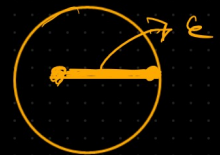
→ DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a popular clustering algorithm used in machine learning, particularly for identifying clusters of points in a dataset

→ It's especially effective when dealing with data that may not be well-suited for traditional clustering methods like K-means, such as data with irregular shapes or varying densities.

DBSCAN → Non-linear

Math Intuition

$\epsilon \rightarrow$ radius



1. Density: DBSCAN operates based on the concept of density. It defines two parameters:

- ϵ (epsilon) → max distance b/w two points

The maximum distance between two points for them to be considered as part of the same neighborhood.

- min Pts → min no of points required to form a dense region.

② Core points, Border points and Noise points

min Pts

not a core point
but lies ϵ neighbourhood
of core points

neither core
nor border
points

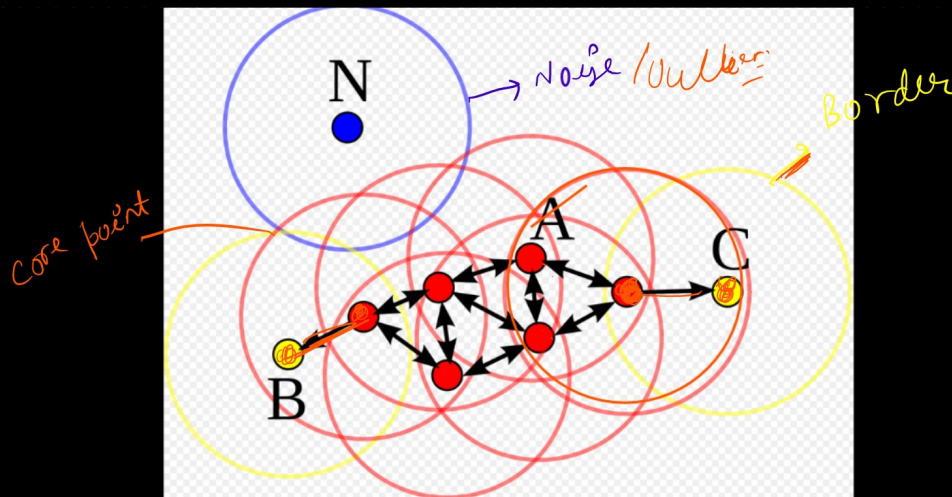
③ Clustering Process

→ DBSCAN Randomly selecting a point from the dataset

→ It then checks if this point is a core point. If it is, it forms a cluster by including all reachable points within ϵ distance.

→ selected point → not a core point

it is assigned to the cluster of some core point within its ϵ -neighborhood.



In this diagram, $\text{minPts} = 4$. Point A and the other red points are core points, because the area surrounding these points in an ϵ radius contain at least 4 points (including the point itself). Because they are all reachable from one another, they form a single cluster. Points B and C are not core points, but are reachable from A (via other core points) and thus belong to the ...

[More details](#)

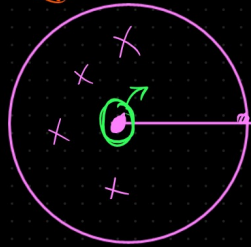
- → core point
- → border point
- → Noise / Outlier

parameter
→ minPts
→ ϵ (epsilon)

core point

No of points within the $\epsilon \geq \text{min_pts}$

core points



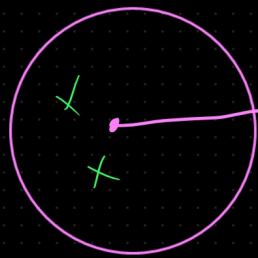
$\text{min_pts} \rightarrow \text{at least } 4$

$$4 \geq$$

(2) Border points

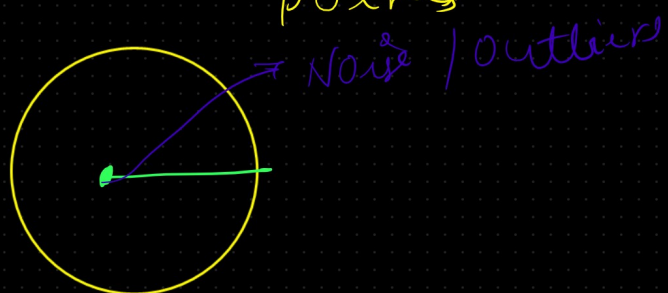
No of a data points within the radius (ϵ) $< \text{min_pts}$

$\text{min_pts} = 4$



Border points

③ Noise \rightarrow Neither Core Nor Border points



Noise / outliers

Supervised ML



Performance matrix



Classification



Accuracy score, Confusion
Classification

Regression



R^2 score, MSE

Silhouette Clustering



quality of cluster



Silhouette score