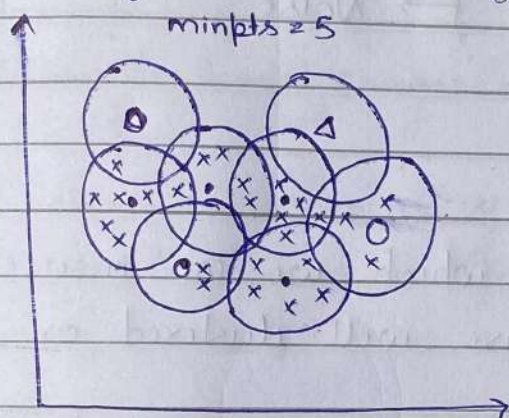


DBScan

→ It is an unsupervised machine learning algorithm which is used to solve clustering problem, especially non-linear type of clustering.



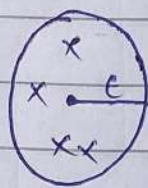
- → core point
- → Border point
- △ → noise / outlier

Hyperparameters

- (i) minpts → minimum points
- (ii) ϵ → radius

* Core Point :

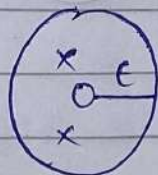
A point is considered as core point when number of points within the radius (ϵ) is greater than or equal to minimum points.



→ minpts = 4 } → core point
 \downarrow
 k=4

* Border Point :

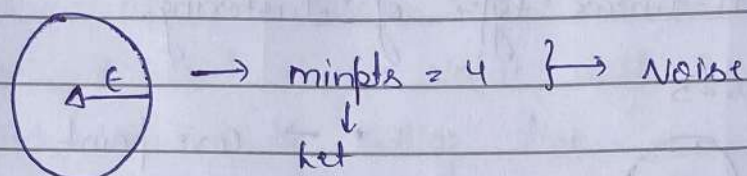
A point is considered as border point when number of points within the radius (ϵ) is less than the minimum points.



→ minpts = 4 } → Border point
 \downarrow
 k=4

* Noise/outliers : { DBScan is robust to outliers }

A point is considered as noise when the point itself present within the radius (ϵ).



* Silhouette clustering :

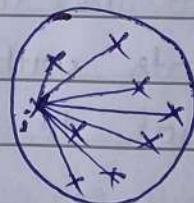
It is a technique through which we can ensure that the formed clusters are well clustered or not.

Steps :

① calculate the mean distance within clusters.

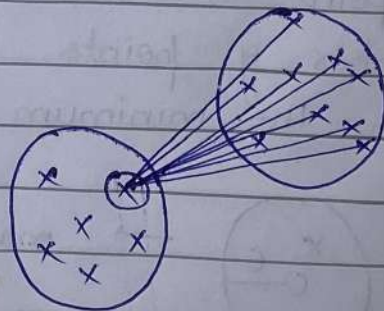
$$a_i = \frac{1}{|C_i| - 1} \sum d(i, j)$$

$\rightarrow C_i$: No. of points.
 d : distance



② calculate the mean distance of data points of nearby clusters.

$$b_i = \min \frac{1}{|C_j|} \sum d(i, j)$$



③ Calculate the Silhouette Score.

$$S = \frac{b_i - a_i}{\max\{a_i, b_i\}} \quad \begin{cases} a_i > b_i \rightarrow \text{Good cluster} \\ a_i < b_i \rightarrow \text{Bad cluster} \end{cases}$$

Silhouette Score ranges $\rightarrow \{0 \text{ to } 1\}$

Note: closer the value of silhouette score, better the clusters are formed.