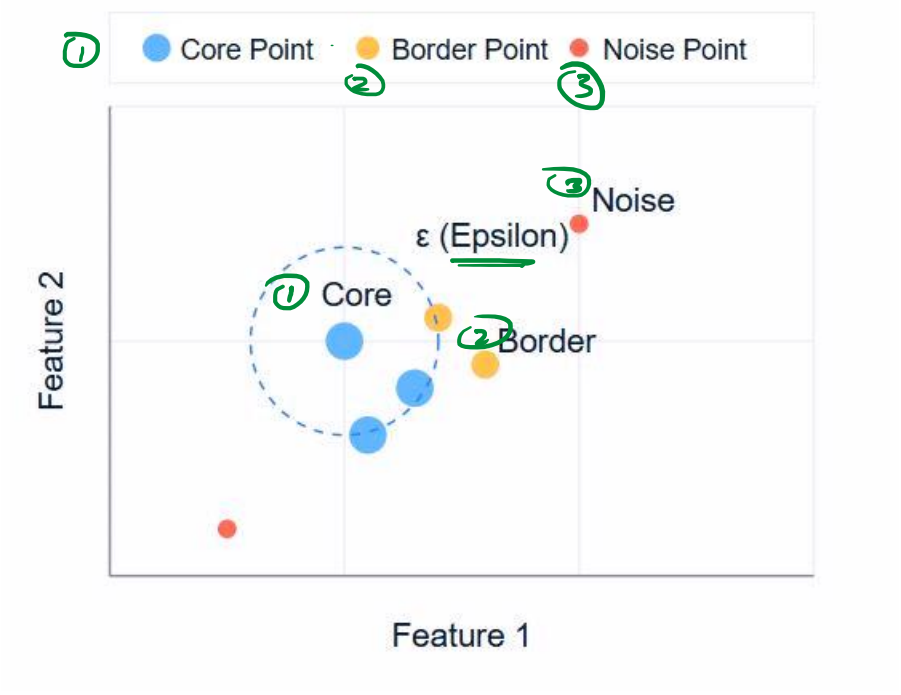
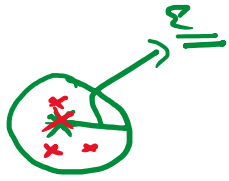
DBSCAN

(Density-Based spatial clustering of Applications with Noise)





1. Core Points → These are points that have at least a minimum number of other points (MinPoints) within a specified distance (ε or epsilon).

2. Border Points → These are points that are

2. [illegible]

within the ε distance of a core point but don't have MinPts neighbors themselves.

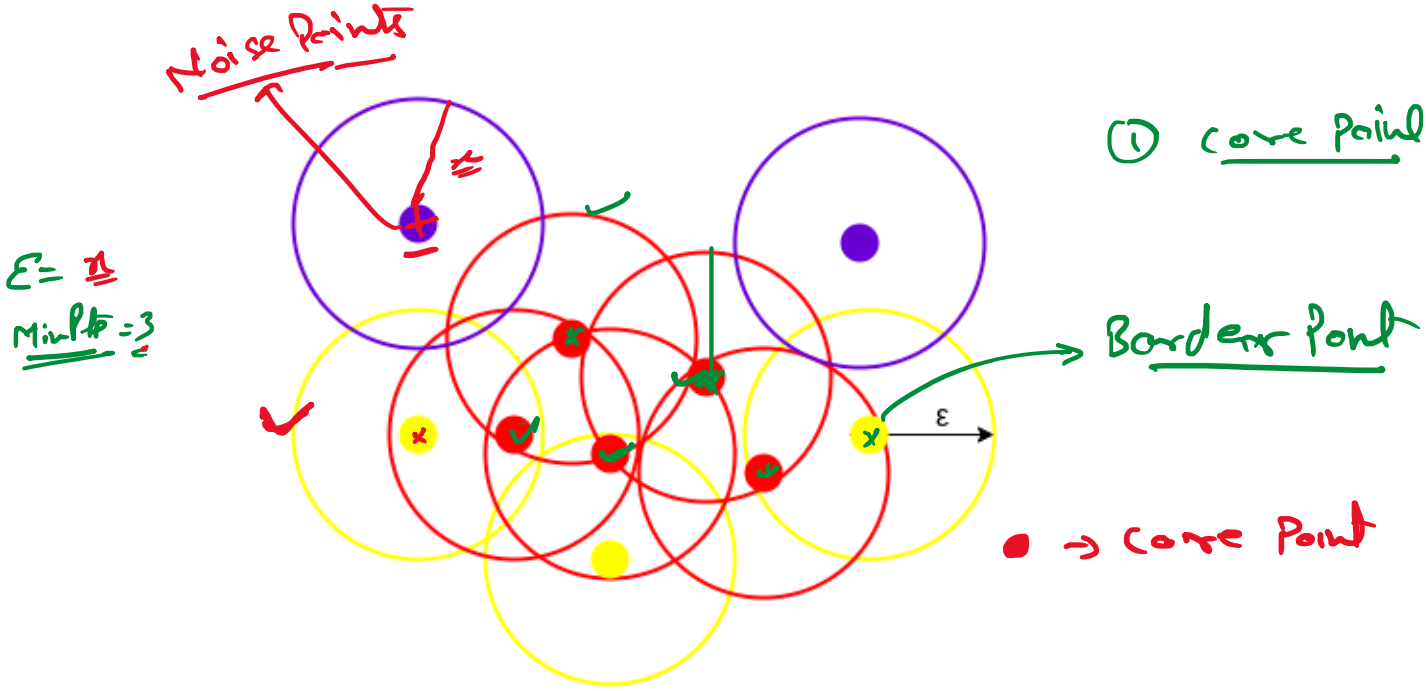
3. Noise Points: These are points that are neither core points nor border points. They're not close enough to any cluster to be included.

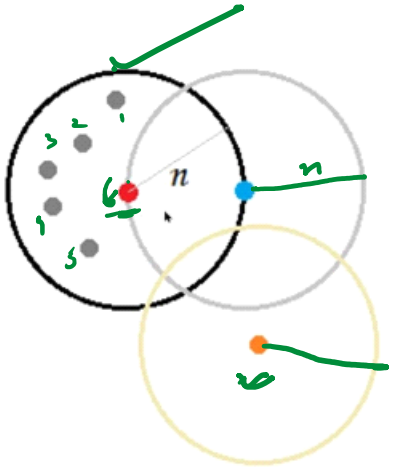
what parameters required DBSCAN

(1) eps (ε) :- This parameter defines the radius of a neighborhood around a data point. Points within this distance are considered neighbors of the central point.

(2) minPts This parameter represents the minimum number of points required within the ε-neighborhood of a point to classify it as a core point. A core point is considered to be dense enough to be part a cluster.

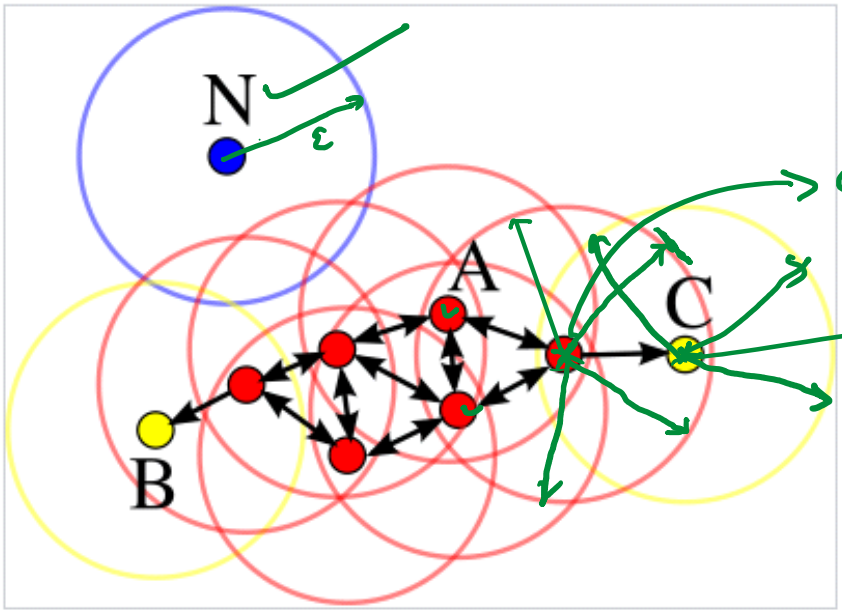
DBSCAN	K-Means
In DBSCAN we need not specify the number of clusters.	K-Means is very sensitive to the number of clusters so it need to specified
Clusters formed in DBSCAN can be of any arbitrary shape.	Clusters formed in K-Means are spherical or convex in shape
DBSCAN can work well with datasets having noise and outliers	K-Means does not work well with outliers data. Outliers can skew the clusters in K-Means to a very large extent.
In DBSCAN two parameters are required for training the Model	In K-Means only one parameter is required is for training the model





DBSCAN CLUSTERING

Screen clipping taken: 19-10-2024 07:34 PM



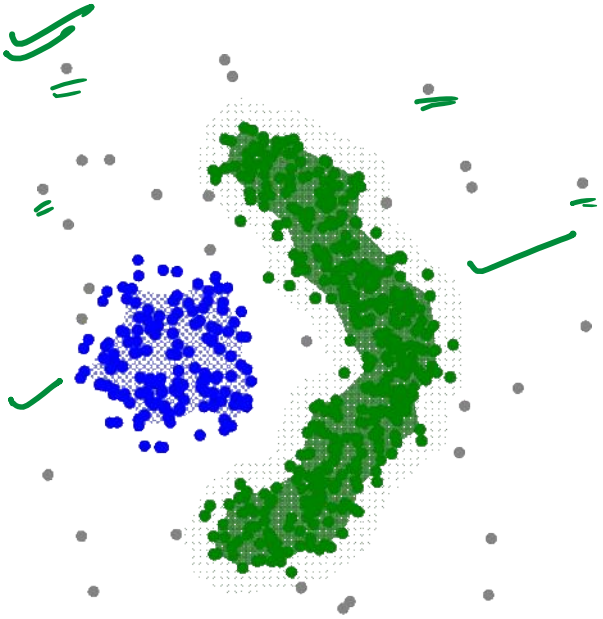
corpoint

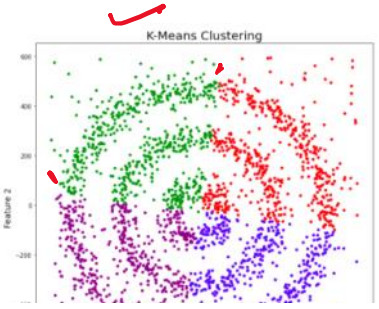
Border Points

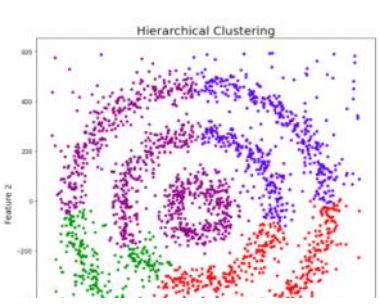
In this diagram, minPts = 4. Point A and the other red points are core points, because the area surrounding these points in an ε radius contain at least 4 points (including the point itself). Because they are all reachable from one another, they form a single cluster. Points B and C are not core points, but are reachable from A (via other core points) and thus belong to the cluster as well. Point N is a noise point that is neither a core point nor directly-reachable.

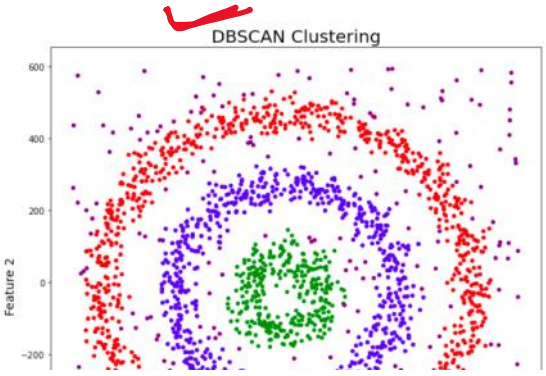
Screen clipping taken: 19-10-2024 07:35 PM

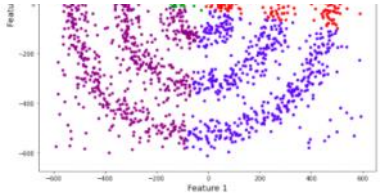
This non linear data points

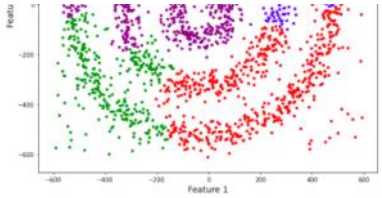


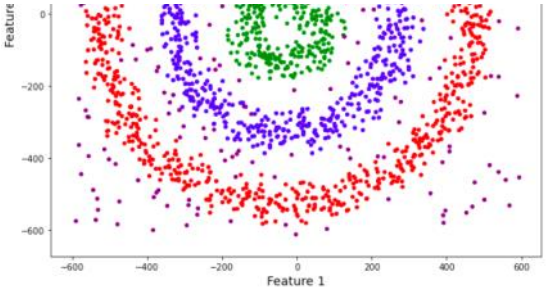












1. https://www.geeksforgeeks.org/dbscan-clustering-in-ml-density-based-clustering/
2. https://medium.com/@sachinsoni600517/clustering-like-a-pro-a-beginners-guide-to-dbscan-6c8274c362c4
3. https://www.analyticsvidhya.com/blog/2020/09/how-dbscan-clustering-works/
4. https://www.datacamp.com/tutorial/dbscan-clustering-algorithm