Density Based Clustering

Necessity

- Clusters with arbitrary shapes

- Clusters of different sizes

- Clusters with different densities

- Finding noise and outliers

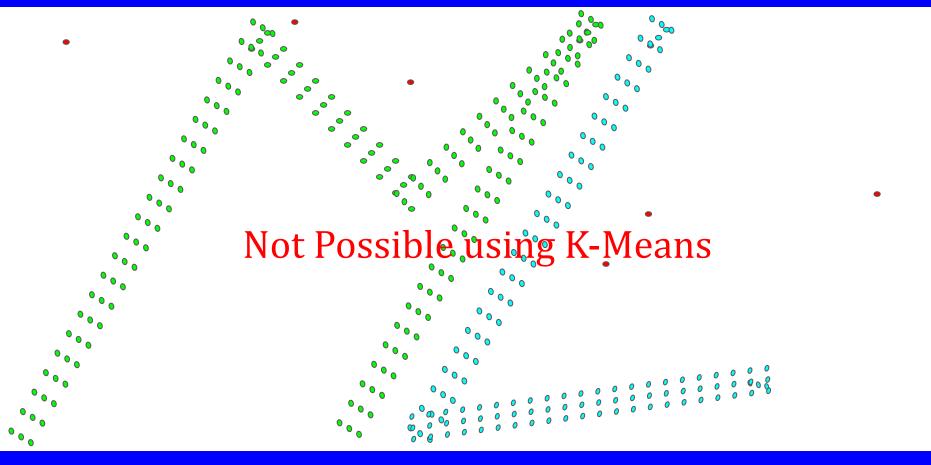# DBSCAN

- Find all the neighbor points within eps and Identify the core points

- For each core point create a new cluster if it doesn't belong to any

- Find recursively all its density connected points and assign them to the same cluster as the core point

- A point a and b are said to be density connected if there exist a point c which has a sufficient number of points in its neighbors and both the points a and b are within the eps distance. This is a chaining process. So, if b is neighbor of c, c is neighbor of d, d is neighbor of e, which in turn is neighbor of a implies that b is neighbor of a.

- Iterate through the remaining unvisited points in the dataset.

- Those points that do not belong to any cluster are noise

Definitions

- ***r***: The radius of the neighborhoods around a data point

- ***minPts***: The minimum number of data points need to be a cluster

- ***Core Points***: A data point p is a core point if it has at least ***minPts*** neighbors

- ***Border Points***: A data point q is a border point if it is not a core point but a neighbor of core point

- ***Outlier***: A data point o is an outlier if it is neither a core point nor a border point

Not Possible using K-Means

Thank You

References

- https://blog.dominodatalab.com/topology-and-density-based-clustering/
- https://www.kdnuggets.com/2020/02/understanding-density-based-clustering.html
- https://www.geeksforgeeks.org/dbscan-clustering-in-ml-density-based-clustering/
-