Unsupervised Learning

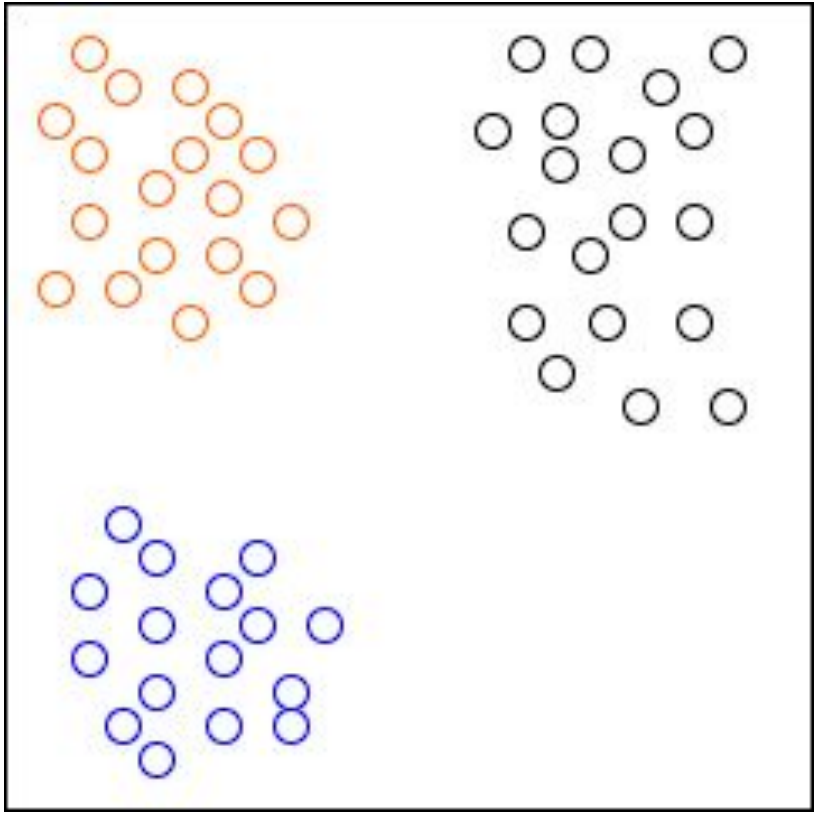
Index

- Introduction To DBSCAN
- Working of DBSCAN

Introduction To DBSCAN

1. DBSCAN stands for **Density-Based Spatial Clustering of Applications** with Noise.
2. It also does not require the number of clusters to be told beforehand, unlike K-Means, where we have to specify the number of centroids.
3. K-Means (distance between points)
4. DBSCAN (distance between nearest points)
5. The key idea is that for each point of a cluster, the neighborhood of a given radius has to contain at least a minimum number of points.

Introduction To DBSCAN



Introduction To DBSCAN

1. Partitioning methods K-means and hierarchical clustering work for finding spherical-shaped clusters or convex clusters.
2. But several data may contain irregularities, like –
   a. i) Clusters can be of convex shape or mixed manner it is difficult to separable out.
   b. ii) Data may contain noise.
3. This algorithm does it by identifying different clusters in the dataset and connects the areas of high densities into clusters.

Working of DBSCAN

1. DBSCAN requires only two parameters:
    a. epsilon and minPoints.
        i. Epsilon is the radius of the circle to be created around each data point to check the density.
        ii. minPoints is the minimum number of data points required inside that circle for that data point to be classified as a Core point.

Working of DBSCAN

1. Epsilon :-
   a. If the distance between two points is lower or equal to ‘eps’ then they are considered as neighbors.
   b. If it is chosen very large then the clusters will merge and majority of the data points will be in the same clusters.

Working of DBSCAN

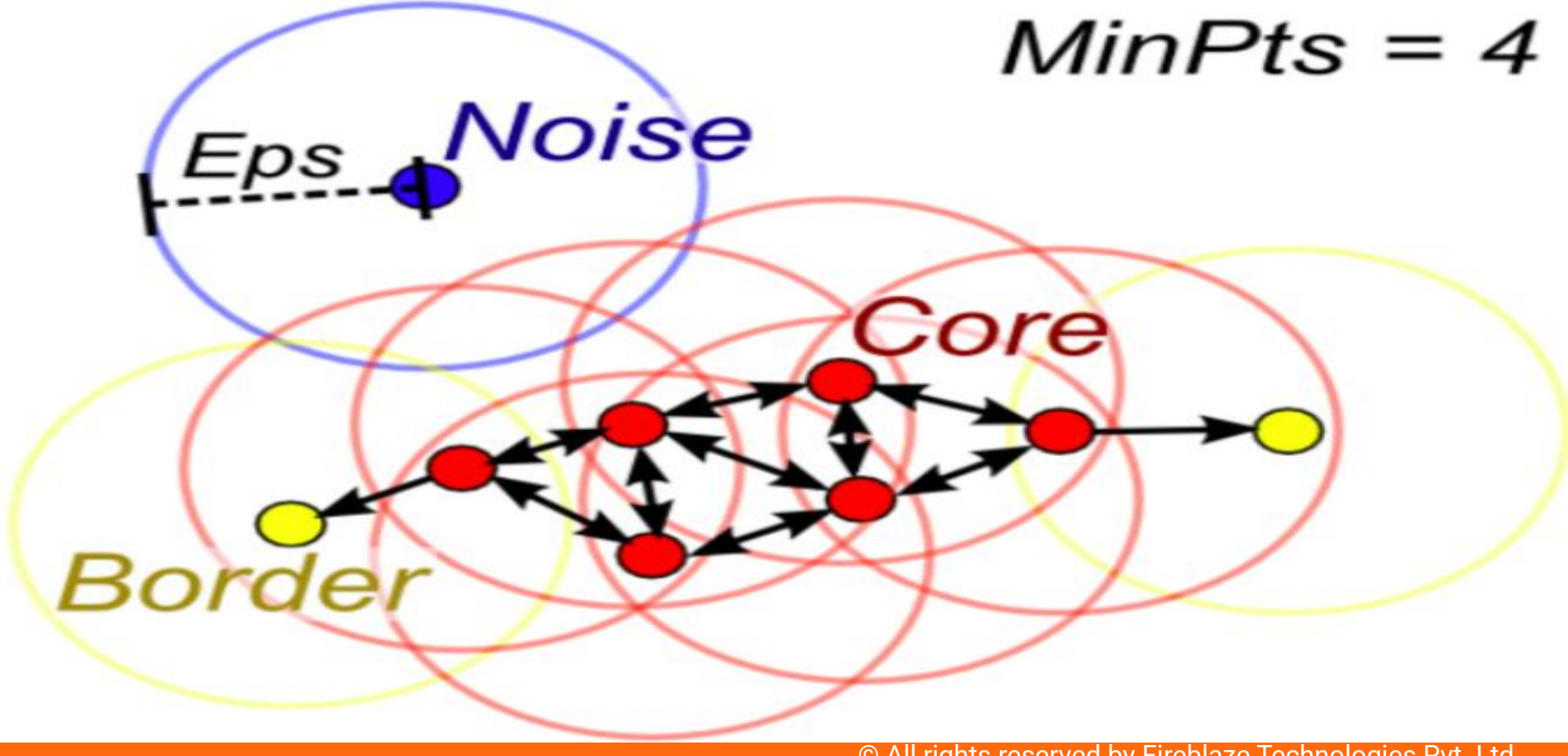
1. minPoints :-
   a. Minimum number of neighbors (data points) within **eps** radius.
   b. Larger the dataset, the larger value of MinPts must be chosen.
   c. The minimum value of MinPts must be chosen at least 3.

Working of DBSCAN

In this algorithm, we have 3 types of data points. :-

a. Core Point: A point is a core point if it has more than MinPts points within eps.
b. Border Point: A point which has fewer than MinPts within eps but it is in the neighborhood of a core point.
c. Noise or outlier: A point which is not a core point or border point.

Working of DBSCAN



Thank you