DBScan

- It is a Density based clustering (Density-based spatial clustering of applications with noise)
- In density based clustering we partition points into dense regions separated by not-so-dense regions.
- DBScan works well even if your data contains non-convex clusters or noise.
- In DBSCAN, there are no centroids, and clusters are formed by linking nearby points to one another.
- **Important things in DBScan**
- **Parameters**
- Eps
- Minpoints
- **Types of data points**
- Core point
- Border point
- Noise or Outlier

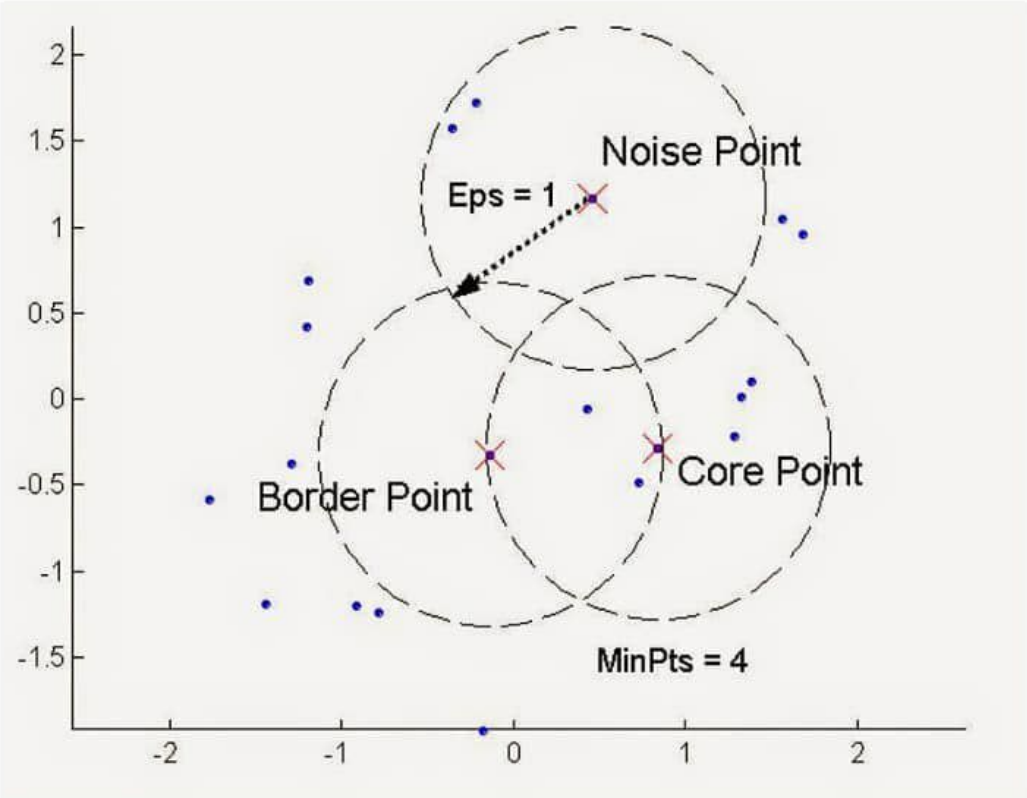
DBScan

- **eps:** It defines the neighborhood around a data point i.e. if the distance between two points is lower or equal to 'eps' then they are considered as neighbors.
- If eps is small, then large no of points will be left out as noise
- If eps is large, then most of thepoints will be in same cluster
- We can find the best ps value using k-distance graph.

DBScan

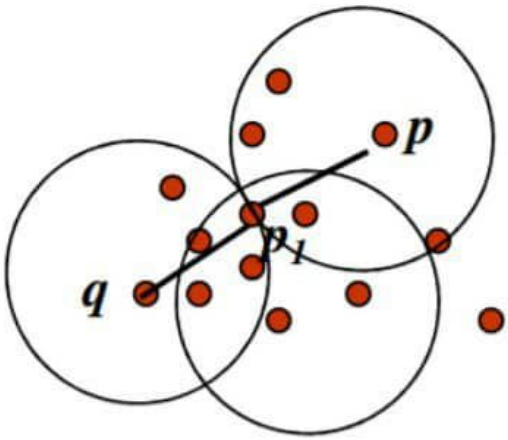
- **MinPoints:** Minimum number of data points within eps radius.
- Larger the dataset, the larger value of MinPts must be chosen.
- As a general rule, the minimum MinPts can be derived from the number of dimensions D in the dataset as,
  MinPts >= D+1.
- Or MinPts = 2*D

DBScan

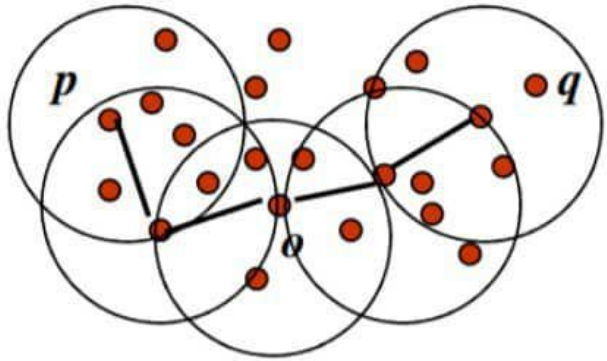


- **Core Point:** A point is a core point if it has more than MinPts points within eps.
- **Border Point:** A point which has fewer than MinPts within eps but it is in the neighborhood of a core point.
- **Noise or outlier:** A point which is not a core point or border point.

DBScan



Density edge



Density connected

- **Density connected points**
- **Density edge**: We place an edge between two core points q and p if they are within distance Eps.
- **Density-connected**: A point p is density-connected to a point q if there is a path of edges from p to q.

DBScan

- **steps**
- Label points as core, border and noise
- Eliminate noise points
- For every core point p that has not been assigned to a cluster
- Create a new cluster with the point p and all the points that are density-connected to p.
- Assign border points to the cluster of the closest core point

DBScan

- **Pros**
- Resistant to Noise
- Can handle clusters of different shapes and sizes
- **Cons**
- Varying densities
- High-dimensional data
- sensitive to parameters