Density Based Clustering

DBSCAN

- DBSCAN is a density-based algorithm.
- DBSCAN requires two parameters: epsilon (Eps) and minimum points (MinPts).
- It starts with an arbitrary starting point that has not been visited .
- It then finds all the neighbour points within distance Eps of the starting point.
- If the number of neighbours is greater than or equal to MinPts, a cluster is formed.
- The starting point and its neighbours are added to this cluster and the starting point is marked as visited.
- The algorithm then repeats the evaluation process for all the neighbours recursively.

DBSCAN

- If the number of neighbours is less than MinPts, the point is marked as noise.
- If a cluster is fully expanded (all points within reach are visited) then the algorithm proceeds to iterate through the remaining unvisited points in the dataset.

Major features:

- Discover clusters of arbitrary shape
- Handle noise
- One scan
- Need density parameters

Basic concept:

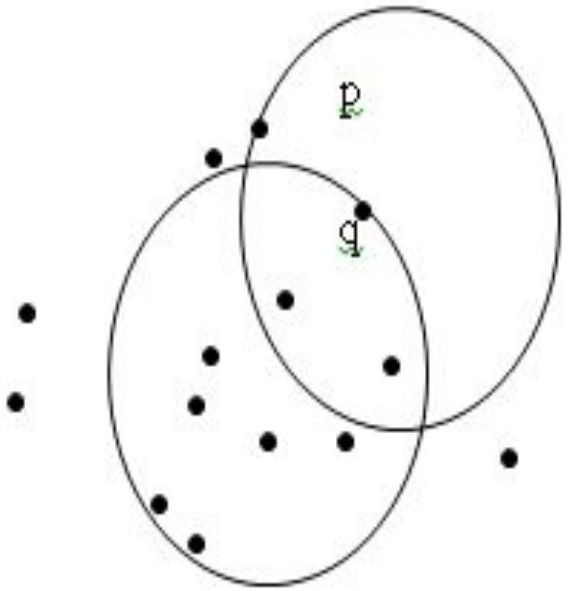
For any cluster we have:

- A central point (p) i.e. core point
- A distance from the core point(Eps)
- Minimum number of points within the specified distance (MinPts)

DBSCAN Algorithm

1. Create a graph whose nodes are the points to be clustered
2. For each core-point c create an edge from c to every point p in the ?-neighborhood of c
3. Set N to the nodes of the graph;
4. If N does not contain any core points terminate
5. Pick a core point c in N
6. Let X be the set of nodes that can be reached from c by going forward;
   a. create a cluster containing X ∪ ∪ {c}
   b. N=N/(X ∪ ∪ {c})
7. Continue with step 4

DBSCAN Algorithm



MinPts= 5

Eps= 1cm

Density-Based Clustering Methods

- Clustering based on density (local cluster criterion), such as density-connected points or based on an explicitly constructed density function
- Major features:
  - Discover clusters of arbitrary shape
  - Handle noise
  - One scan
  - Need density parameters
- Several interesting studies:
  - DBSCAN: Ester, et al. (KDD'96)
  - DENCLUE: Hinneburg & D. Keim (KDD'98/2006)
  - OPTICS: Ankerst, et al (SIGMOD'99).
  - CLIQUE: Agrawal, et al. (SIGMOD'98)

Density-Based Clustering Methods

- DBSCAN is a density-based algorithm.
  - Density = number of points within a specified radius r (Eps)
  - A point is a core point if it has more than a specified number of points (MinPts) within Eps
    - These are points that are at the interior of a cluster
  - A border point has fewer than MinPts within Eps, but is in the neighborhood of a core point
  - A noise point is any point that is not a core point or a border point.

DBSCAN: Core, Border, and Noise Points