Density-Based Clustering Methods

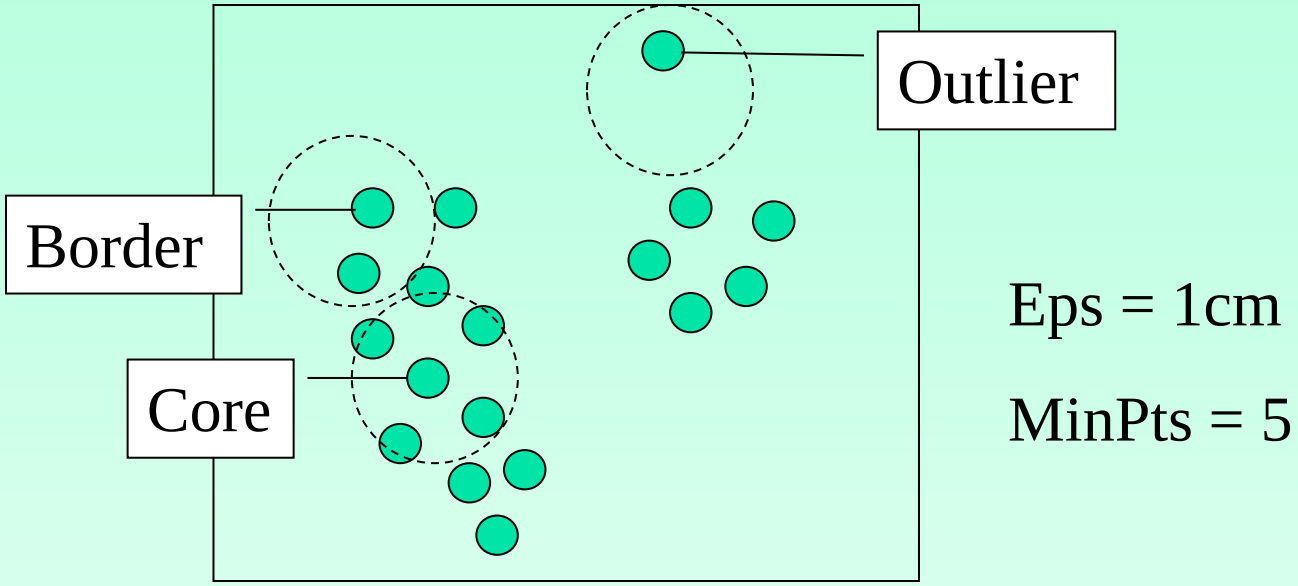
- Clustering based on density (local cluster criterion), such as density-connected points
- Major features:
  - Discover clusters of arbitrary shape
  - Handle noise
  - Need density parameters as termination condition
- Several interesting studies:
  - DBSCAN
  - OPTICS

DBSCAN: Density-Based Spatial Clustering of Applications with Noise

- Relies on a *density-based* notion of cluster: A *cluster* is defined as a maximal set of density-connected points
- Discovers clusters of arbitrary shape in spatial databases with noise

Density based method

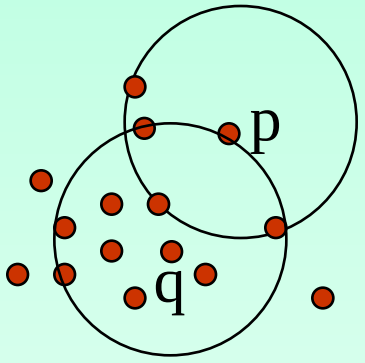
- **Core Point:** A point is a core point, if number of points surrounding it is greater than minimum points for a given radius.
- **Border Point:** It is not a core point but falls within a neighborhood of core point/several core points.
- **Noise Point:** Point that is neither core point nor border point.



Density-Based Clustering: Basic Concepts

- Two parameters:
  - *Eps*: Maximum radius of the neighbourhood
  - *MinPts*: Minimum number of points in an Eps-neighbourhood of that point
- $N_{Eps}(p)$: {q belongs to D | dist(p,q) ≤ Eps}
- Directly density-reachable: A point *p* is directly density-reachable from a point *q* w.r.t. *Eps*, *MinPts* if
  - *p* belongs to $N_{Eps}(q)$
  - core point condition:

$$|N_{Eps}(q)| \geq MinPts$$



MinPts = 5

Eps = 1 cm

DBSCAN: The Algorithm

- Arbitrary select a point *p*
- Retrieve all points density-reachable from *p* w.r.t. *Eps* and *MinPts*
- If *p* is a core point, a cluster is formed
- Any border point that is close enough to a core point is put in the same cluster as the core point.
- Discard noise point.
- Continue the process until all of the points have been processed

Pros and Cons

- **Pros**
  - Not sensitive to noise
  - It can find clusters of arbitrary shape and size.

- **Cons**
  - For high dimensional data, density is difficult to define.
  - Very expensive because of computation.