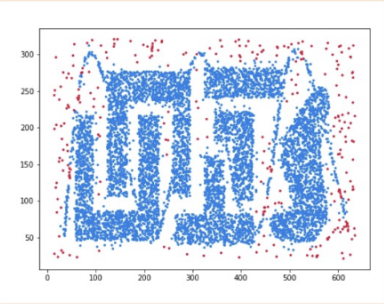
✱ Density-based Clustering

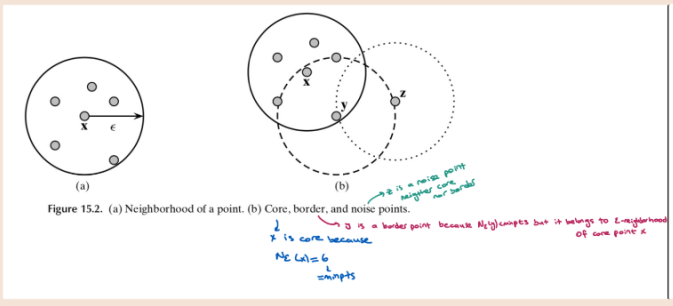
- It can mine non-convex clusters. (Representative-based or other methods have trouble with non-convex clusters)

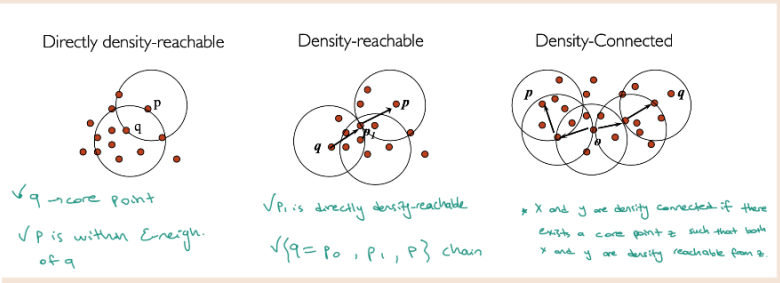


- Clustering based on density
- Major features;
  - Discover clusters of arbitrary shape
  - Handle noise
  - One scan
  - Need density parameters as termination condition

>> DBSCAN

- Density corresponds to have at least minpts points within specific radius ε
- Core point contains at least minpts objects in its ε-neighborhood
- Border point has fewer than minpts within ε, but is in the neighborhood of a core point
- Noise point is any point that is not a core point nor a noise point





- Directly density reachable: x is directly density-reachable from y if x is within the ε-neighborhood of y and y is a core point.
- Density reachable: X is density-reachable from object y if there is a chain $x_1, \ldots, x_n$ where $x_1 = x$ and $x_n = y$ such that x_{i+1} is directly density reachable from x_i
- Density Connected: p is density connected to q with respect to ε and minpts if there is "o" such that both p and q are density reachable from o
- Density-Based Cluster: Maximal set of density connected points.

- Pseudocode:

Algorithm 15.1: Density-based Clustering Algorithm

```
DBSCAN (D, ε, minpts):
1  Core ← ∅
2  foreach x_i ∈ D do // Find the core points
3      Compute N_ε(x_i)
4      id(x_i) ← ∅ // cluster id for x_i
5      if N_ε(x_i) ≥ minpts then Core ← Core ∪ {x_i}
6  k ← 0 // cluster id
7  foreach x_i ∈ Core, such that id(x_i) = ∅ do
8      k ← k + 1
9      id(x_i) ← k // assign x_i to cluster id k
10     DENSITYCONNECTED (x_i, k)
11 C ← {C_i}_{i=1}^k, where C_i ← {x ∈ D | id(x) = i}
12 Noise ← {x ∈ D | id(x) = ∅}
13 Border ← D \ {Core ∪ Noise}
14 return C, Core, Border, Noise

DENSITYCONNECTED (x, k):
15 foreach y ∈ N_ε(x) do
16     id(y) ← k // assign y to cluster id k
17     if y ∈ Core then DENSITYCONNECTED (y, k)
```

- DBSCAN fails when applied to data of varying density
- HDBScan convert DBSCAN into a hierarchical clustering algorithm