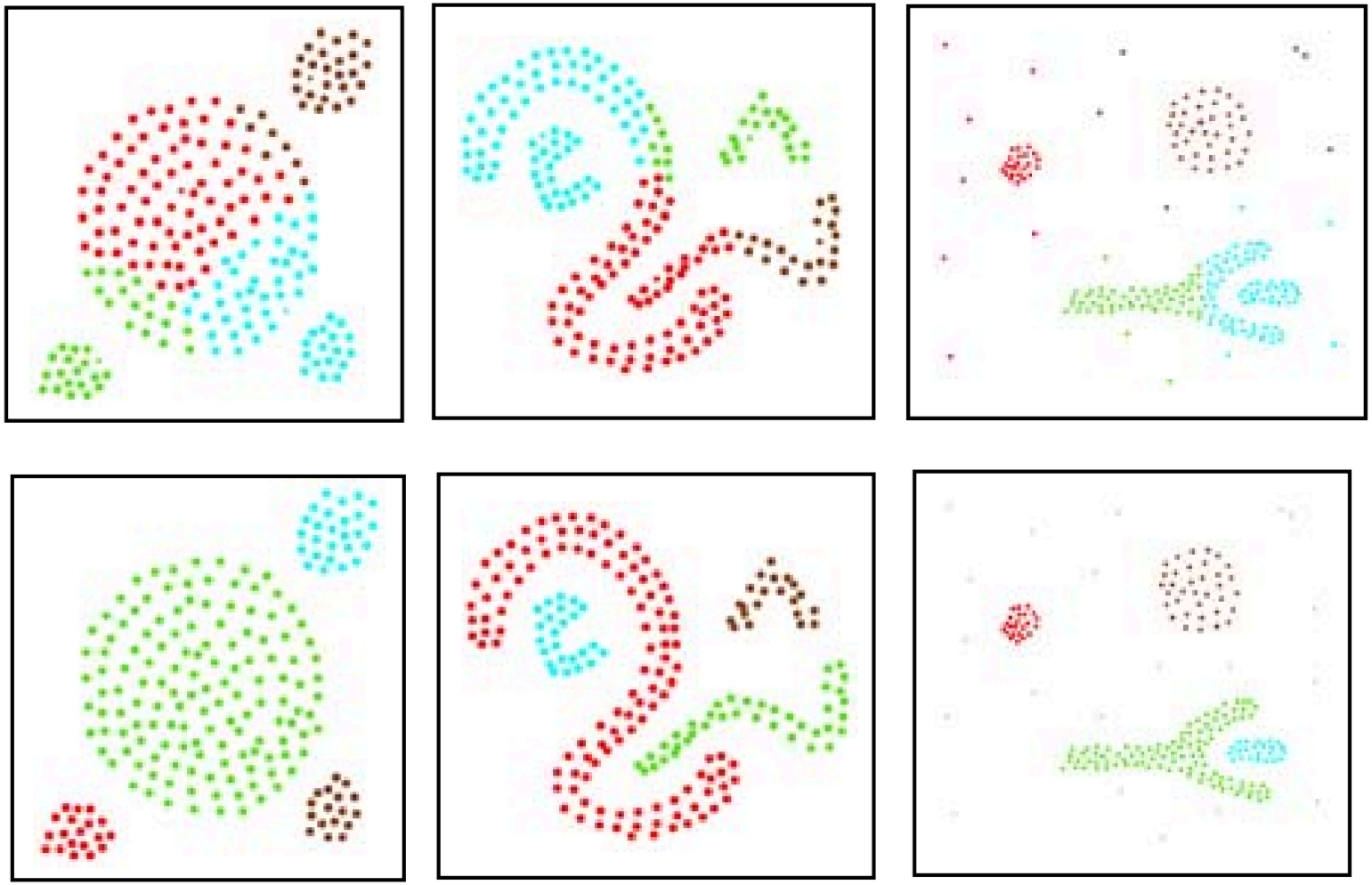
CLUSTERING

k-means

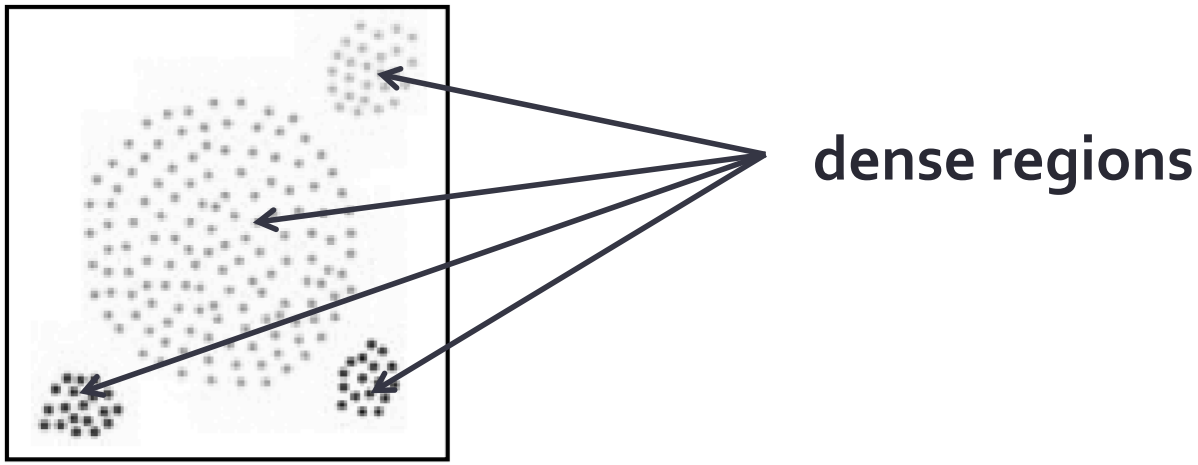
DBSCAN

Why Density-Based Clustering ?



Density-Based Clustering

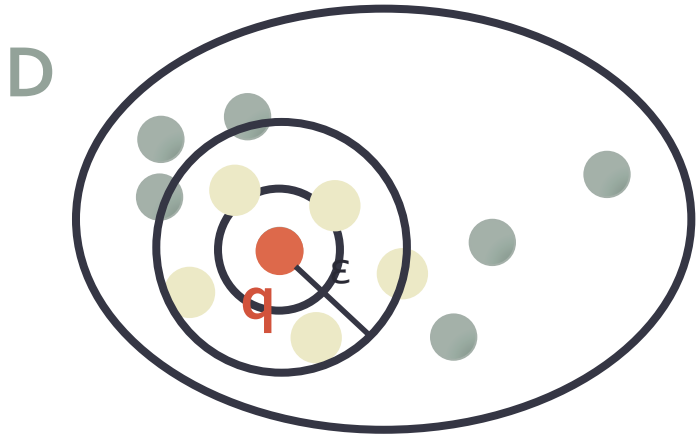
- *Basic Idea*: Clusters are dense regions in the data space, separated by regions of lower object density



- Different density-based clustering approaches exist

DBSCAN

- **D**ensity **B**ased **S**patial **C**lustering of **A**pplications with **N**oise
- The basic idea behind a cluster:
  - For any point in a cluster, the local point density around that point has to exceed some threshold
  - The set of points from one cluster is spatially connected
- What is point density ?



ε-radius defining the neighborhood of a point

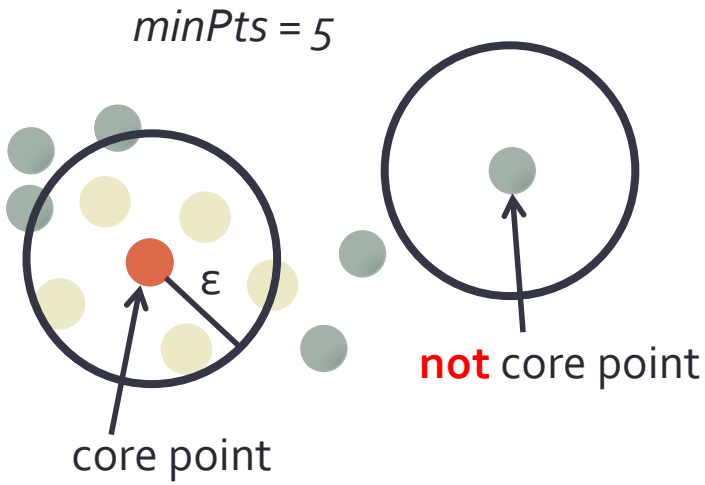
$$N_\varepsilon(q) := \{ d \in D \mid dist(d,q) \leq \varepsilon \}$$

min. number of points in a dense region

$$|N_\varepsilon(q)| \geq minPts \rightarrow \textbf{dense region}$$

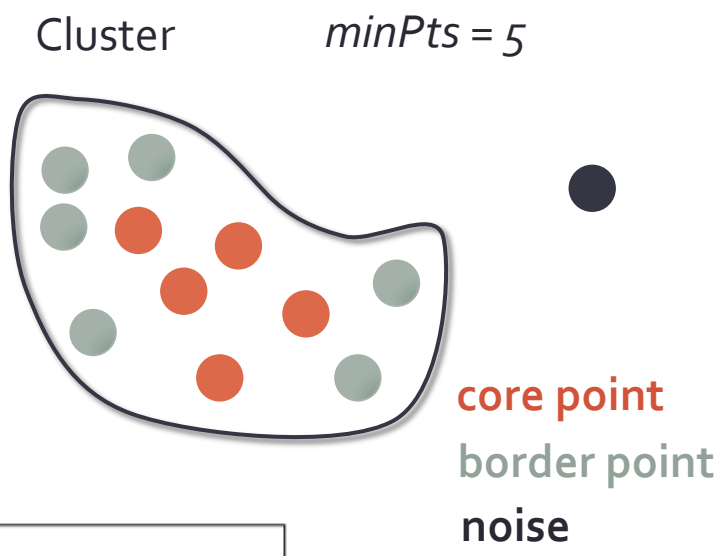
DBSCAN : Terminology (I)

- q is a **core point** if $| N_\varepsilon(q) | \geq$ minPts
- p is **directly density-reachable** from q if
  - $p \in N_\varepsilon(q)$ and
  - q is a core point
- **density-reachable**: transitive closure of *directly* density-reachable
- p is **density-connected** to a point q if there is a point o such that both, p and q are density-reachable from o



DBSCAN : Terminology (II)

- A **cluster** S is a set of points o∈D that satisfies:
  - *Maximality*: if p is in S and q is density-reachable from p then q is in S
  - *Connectivity*: each object in S is density-connected to all other objects in S




```
for each o ∈ D do
    if o is not yet classified then
        if o is a core-object then
            collect all objects density-reachable from o
            and assign them to a new cluster.
        else
            assign o to NOISE
```

Example

- $\varepsilon = 2.0$
- $MinPts = 3$