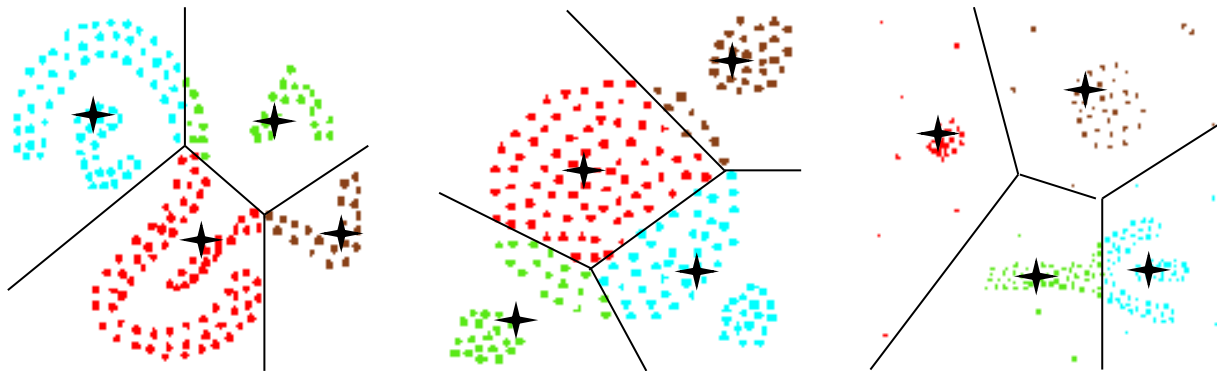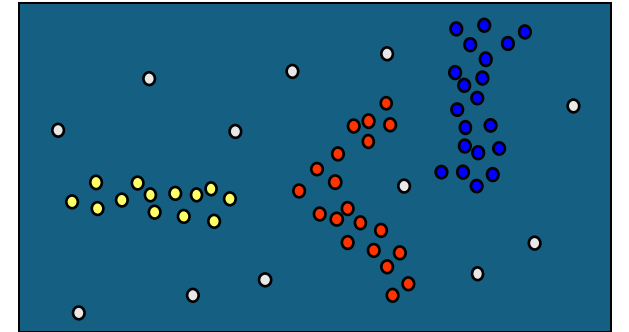# Density Based Clustering

DBSCAN (Density based spatial clustering of applications with noise)

- **Most partitioning (clustering) methods cluster objects based on the distance between objects**

- **Such methods can find only spherical-shaped clusters and encounter difficulty at discovering of arbitrary shapes**
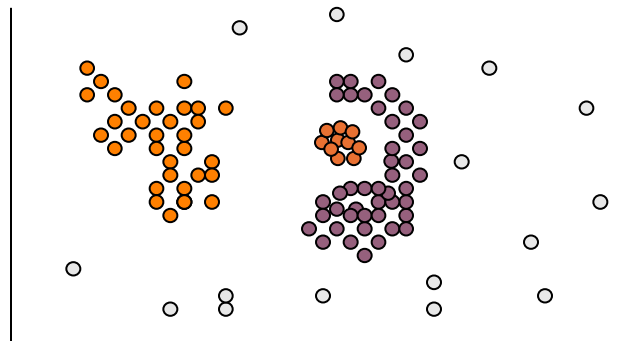

Results of a *k*-means algorithm for *k*=4

Density Based Clustering

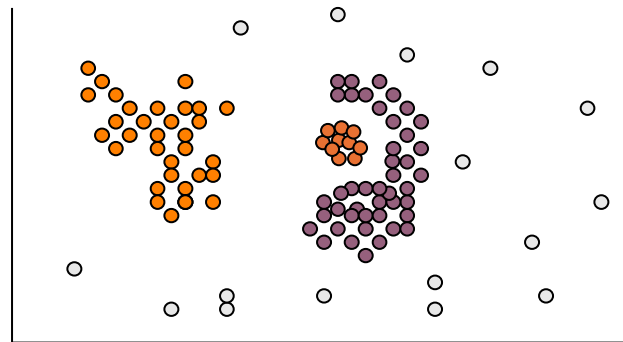## *DBSCAN (Density based spatial clustering of applications with noise)*

- Other methods have been developed based on the notion of density

- Their general idea is to continue growing the given cluster as long as the density (number of data points) in the neighbourhood exceeds some threshold

Density Based Clustering

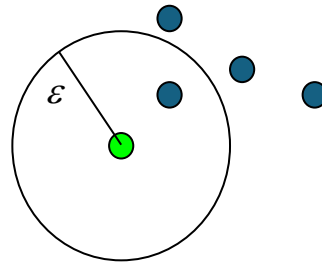## *DBSCAN (Density based spatial clustering of applications with noise)*

- That is, for each data point within a given cluster, the neighbourhood of a given radius has to contain at least a minimum number of points

- Such methods can be used to filter out noise (outliers), and discover clusters of arbitrary shape
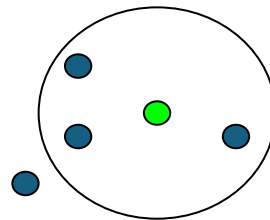
Density Based Clustering

DBSCAN (Density based spatial clustering of applications with noise)

- **The neighbourhood within a radius ε of a given object (data point) is called the ε neighbourhood of the object.**
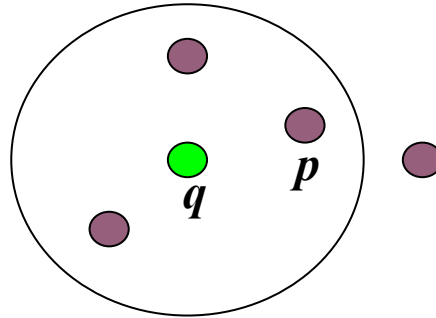

- **If the ε neighbourhood of an object contains at least a minimum number, *MinPts*, of objects, then the object is called a _core_ object.**

Density Based Clustering

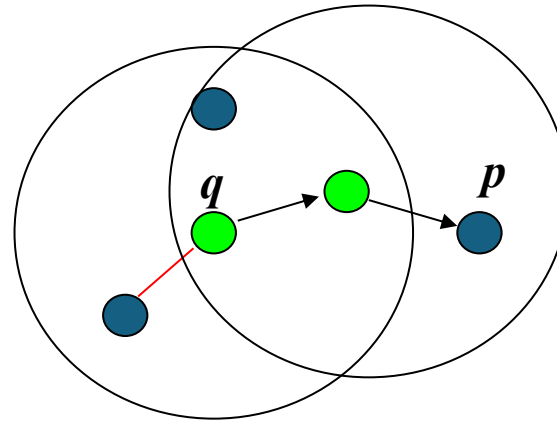## *DBSCAN (Density based spatial clustering of applications with noise)*

- **We say that an object *p*, is directly density-reachable from a core object *q* if *p* is within the ε neighbourhood of *q***

Density Based Clustering

DBSCAN (Density based spatial clustering of applications with noise)
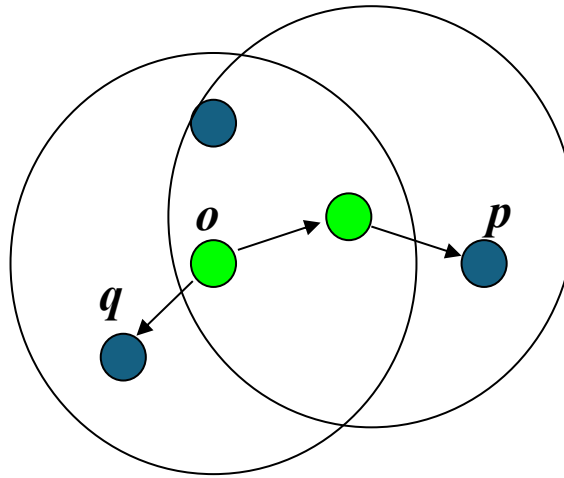
- An object p is density-reachable from object q, if there is a chain of objects $p_1, \ldots, p_n$; where $p_1 = q$ and $p_n = p$ such that p_{i+1} is directly density reachable from p_i

Density Based Clustering
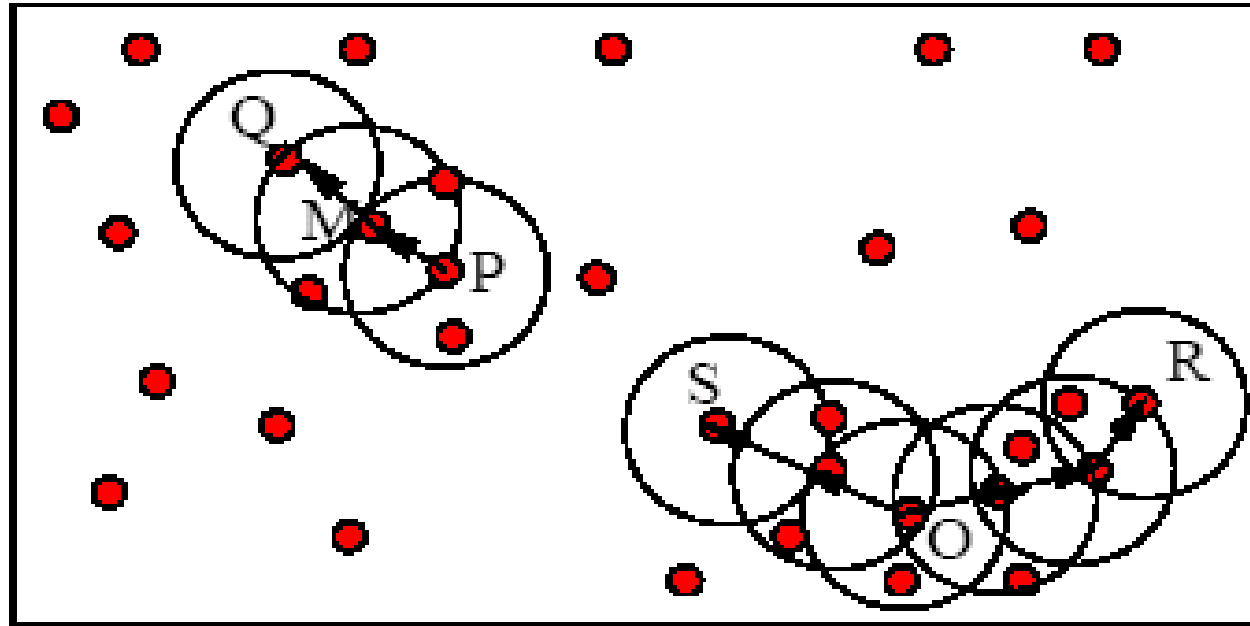
DBSCAN (Density based spatial clustering of applications with noise)

- An object *p* is **density-connected** to object *q*, if there is an object *o* such that both *p* and *q* are density reachable from *o*

Density Based Clustering

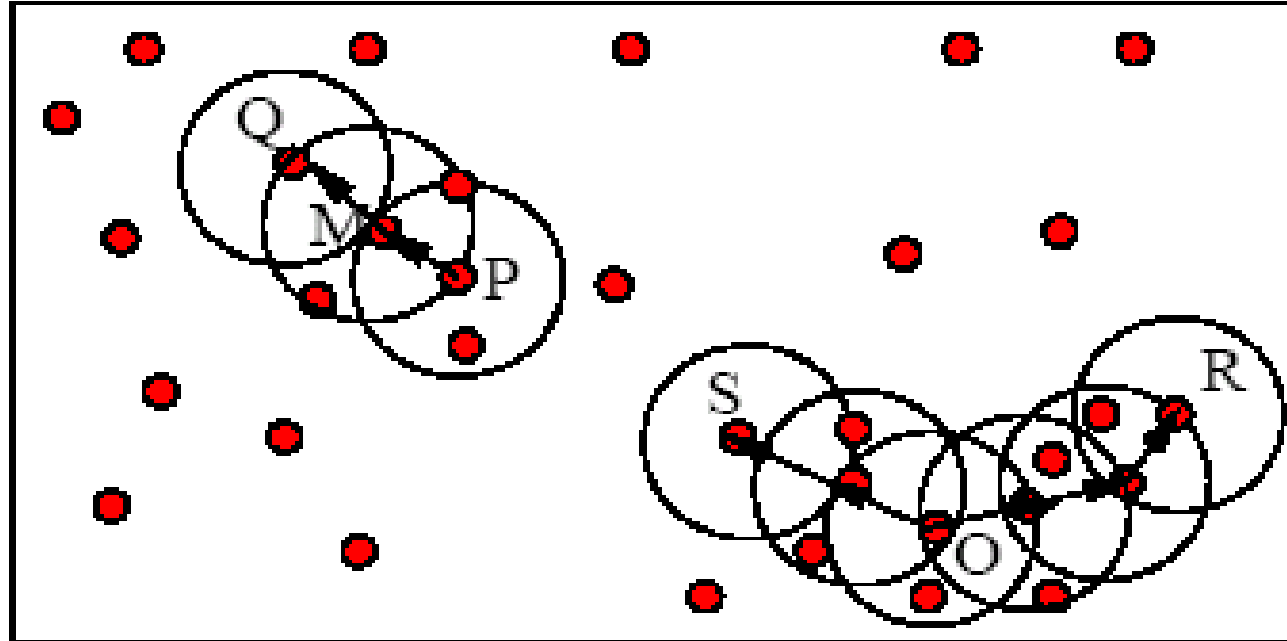## *DBSCAN (Density based spatial clustering of applications with noise)*

- **M, P, O and R are core objects, if *MinPts* = 3**
- **P and Q are directly density reachable from M and M is directly density reachable from P**

Density Based Clustering

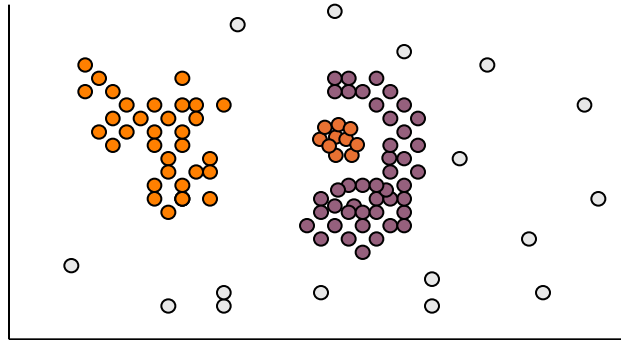## *DBSCAN (Density based spatial clustering of applications with noise)*

- **Q is density reachable** from P, but P is not density reachable from Q
- S, O, R are all **density connected**

Density Based Clustering

DBSCAN (Density based spatial clustering of applications with noise)

A **cluster** is a set of density-connected objects

**Noise** is the set of objects not contained in
any cluster.

Density Based Clustering

DBSCAN (Density based spatial clustering of applications with noise)

- Arbitrary select a point *p*

- Retrieve all points density-reachable from *p* wrt *Eps* and *MinPts*.

- If *p* is a core point, a cluster is formed.

- If *p* is a border point, no points are density-reachable from *p* and DBSCAN visits the next point of the database.

- Continue the process until all of the points have been processed.
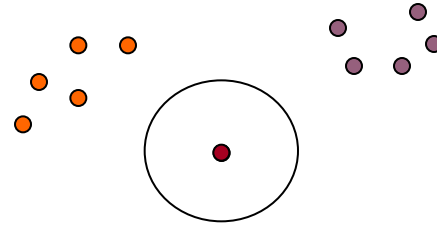
DBSCAN Algorithm: Example
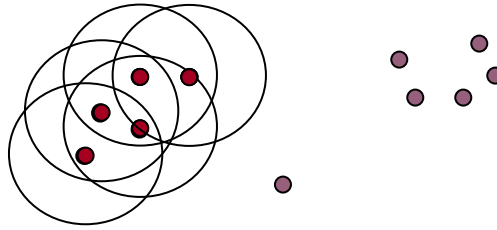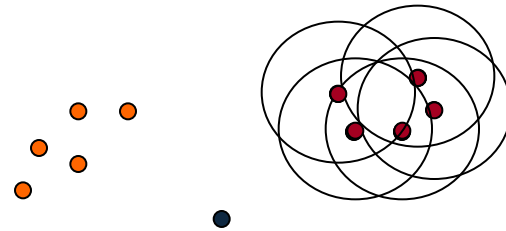
- Parameter
  - ε = 2 cm
  - *MinPts* = 3

DBSCAN Algorithm: Example

- Parameter
    - ε = 2 cm
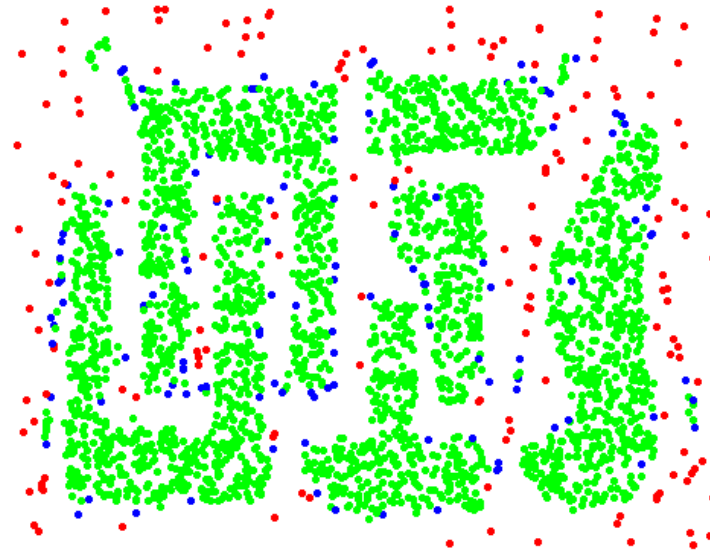   - *MinPts* = 3

DBSCAN Algorithm: Example

- Parameter
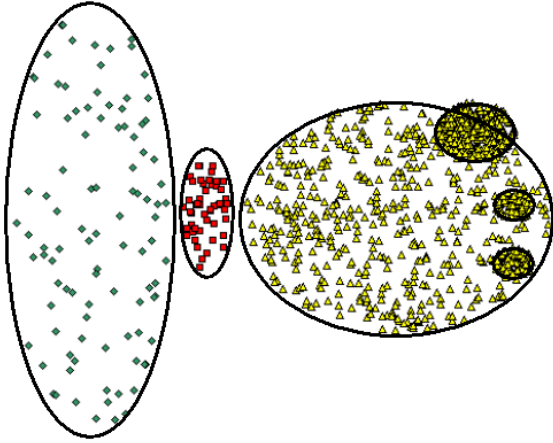   - ε = 2 cm
    - *MinPts* = 3

Example

Original Points

Point types: core, border and outliers
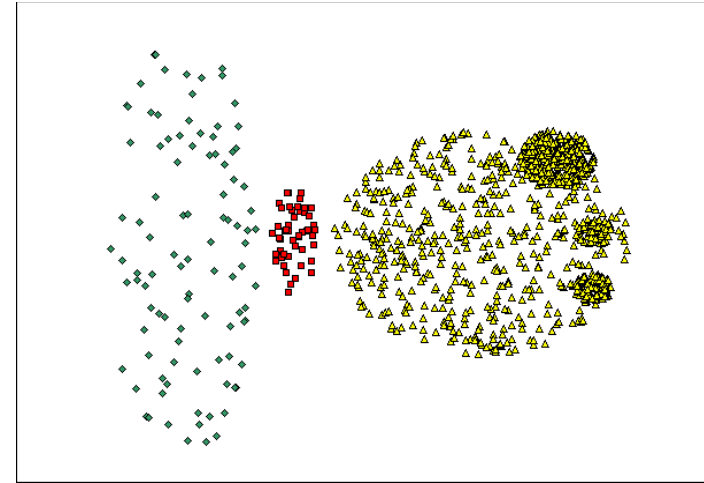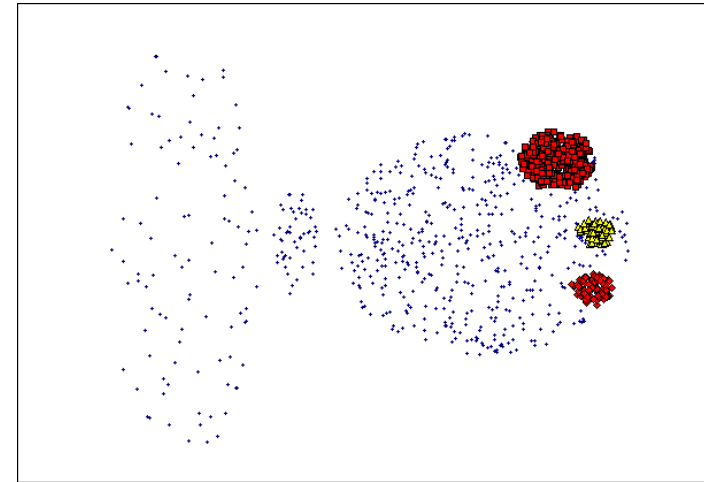
ε = 10, MinPts = 4

When DBSCAN Does NOT Work Well

Original Points

(MinPts=4, Eps=9.92).

(MinPts=4, Eps=9.75)

- **cannot handle Varying densities**

- **sensitive to parameters**