Data Mining

Mean-shift algorithm

16/11/2021

NOVA-IMS

Fernando Lucas Bação

bacao@isegi.unl.pt

http://www.isegi.unl.pt/fbacao

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

1

AGENDA

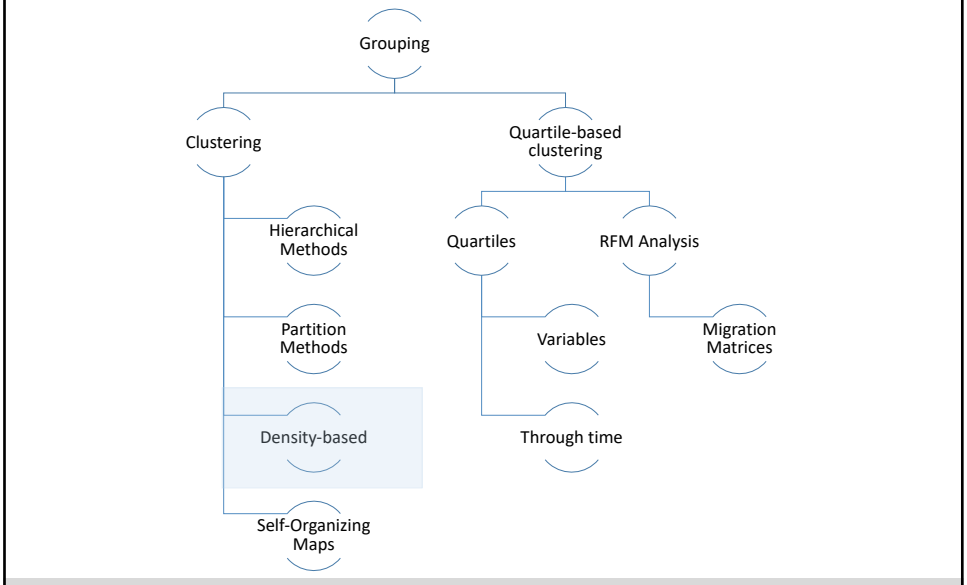
- Cluster analysis
  - Clustering techniques
    - Density-based clustering (mean shift clustering)

2

Clustering

3

Clustering



4

Mean-Shift Clustering

5

- **Characteristics**
  - Mean shift clustering is a sliding-window-based algorithm that attempts to find dense areas of data points.
  - It is a centroid-based algorithm meaning that the goal is to locate the center points of each group/class,
    - Works by updating candidates for center points to be the mean of the points within the sliding-window.
    - These candidate windows are then filtered in a post-processing stage to eliminate near-duplicates, forming the final set of center points and their corresponding groups.

6

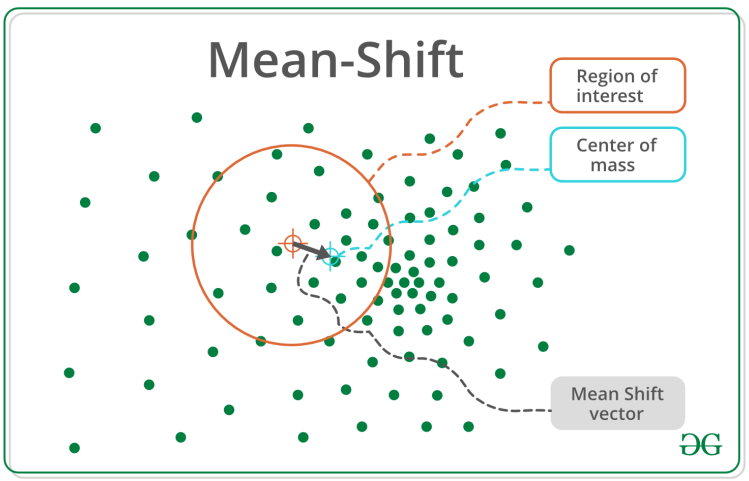
Mean-Shift Clustering

- **Algorithm**

1. Begin with a circular sliding window centered at a point C (randomly selected) and having radius r as the kernel.
2. At every iteration, the sliding window is shifted towards regions of higher density by shifting the center point to the mean of the points within the window (will gradually move towards areas of higher point density).
3. We continue shifting the sliding window according to the mean until there is no direction at which a shift can accommodate more points inside the kernel.
4. This process of steps 1 to 3 is done with many sliding windows until all points lie within a window. When multiple sliding windows overlap the window containing the most points is preserved. The data points are then clustered according to the sliding window in which they reside.

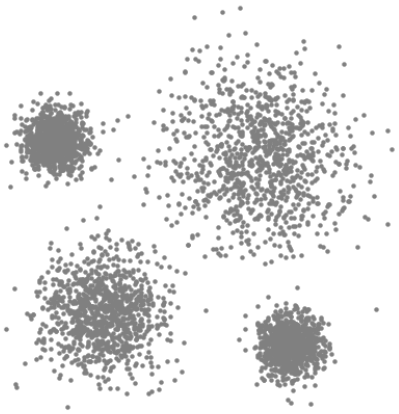
7

Mean-Shift Clustering



8

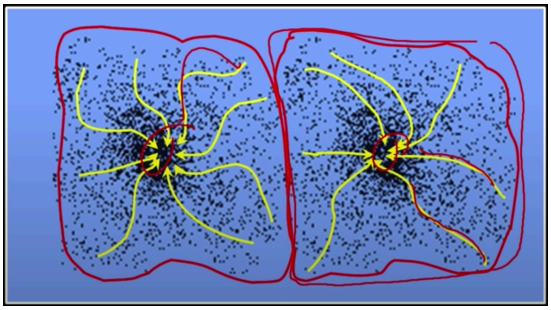
Mean-Shift Clustering



9

Mean-Shift Clustering

- **Characteristics**
  - **Cluster**: all the data points in the attraction basin of a mode
  - **Attraction basin**: the region where all trajectories lead to the same mode



10

11