

Study and visualiser of different clustering methods for density-based clustering

Martin Loginov, Hans Mäesalu, Sven
Aller

Main goals

- Study different density-based clustering methods
 - DBSCAN
 - DCBOR
 - SNN
- Implement a clustering visualisation application

DBSCAN

Density-based spatial clustering of applications with noise

- Based on density based notion of clusters
- Can discover clusters of different shapes
- Has notion of noise
- Do not need to know the number of clusters beforehand

DCBOR

Density Clustering Based on Outlier Removal

- Based on the notion of local density
 - needs to find k-nearest neighbors of each datapoint
 - we use a KD-tree
- Two phases
 - Outlier removal
 - Only 1 input parameter!
 - Clustering
 - cross between DBSCAN and single-link

SNN

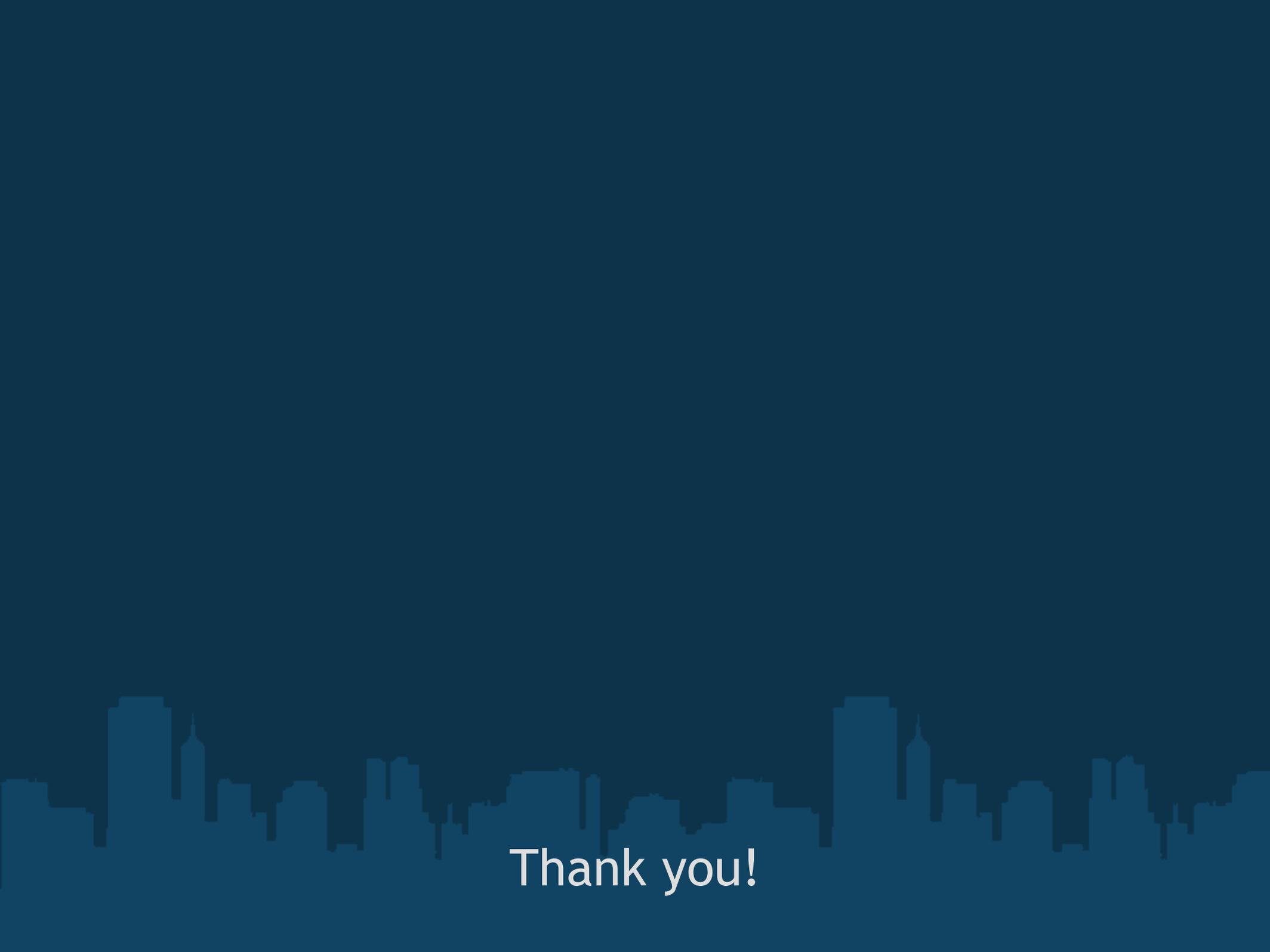
Shared nearest neighbor

- Similarity definition based on the number of shared nearest neighbors
 - If point p_1 is close to point p_2 and they are both close to a set of points, S , then their similarity is equal to the number of points in set S
- Density is defined as the number of points that are similar to a point

Implementation

The screenshot shows a software window titled "Density-based clustering". The interface includes a logo with the text "Density-based clustering" and a small 3D cube icon. Below the logo, there is a control panel with a "Number of points: 1" label and a slider. There are three buttons: "Save", "Open", and "Clear". Below these are three tabs: "DBSCAN", "DCBOR", and "SNN", with "DBSCAN" selected. The "Eps:" parameter is set to 10 and "minPts:" is set to 5. A list box below shows "DBSCAN" selected. The main area is a scatter plot with 3713 points. A central cluster of red points is surrounded by a ring-shaped cluster of green points. Other points are grey. On the right, a panel shows "Elements: 3713", "Clusters: 2", and a "Show cluster" list with "All", "Noise", "1 (154)", and "2 (3265)".

Authors: Martin Loginov, Hans Mäesalu, Sven Aller



Thank you!