Ludwig-Maximilians-Universität München
Institut für Informatik
Prof. Dr. Thomas Seidl
Janina Sontheim, Maximilian Hünemörder

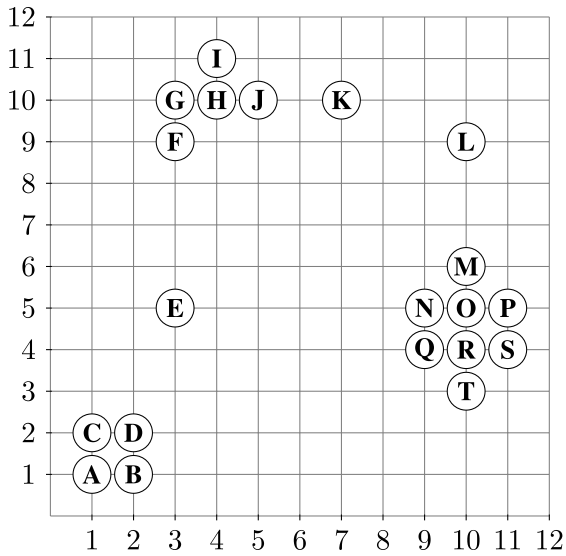
Knowledge Discovery and Data Mining I

WS 2019/20

Exercise 7: DBSCAN, Spectral Clustering

Exercise 7-1 DBSCAN

Given the following data set:



As distance function, use Manhattan Distance:

$$L_1(x, y) = |x_1 - y_1| + |x_2 - y_2|$$

Compute DBSCAN and indicate which points are core points, border points and noise points.

Use the following parameter settings:

- Radius $\varepsilon = 1.1$ and *minPts* $= 2$
- Radius $\varepsilon = 1.1$ and *minPts* $= 3$
- Radius $\varepsilon = 1.1$ and *minPts* $= 4$
- Radius $\varepsilon = 2.1$ and *minPts* $= 4$
- Radius $\varepsilon = 4.1$ and *minPts* $= 5$
- Radius $\varepsilon = 4.1$ and *minPts* $= 4$

Exercise 7-2 Properties of DBSCAN

Discuss the following questions/propositions about DBSCAN:

- Using *minPts* $= 2$, what happens to the border points?
- The result of DBSCAN is deterministic w.r.t. the core and noise points but not w.r.t. the border points.
- A cluster found by DBSCAN cannot consist of less than *minPts* points.
- If the dataset consists of n objects, DBSCAN will evaluate exactly n ϵ-range queries.
- On uniformly distributed data, DBSCAN will usually either assign all points to a single cluster or classify every point as noise. k-means on the other hand will partition the data into approximately equally sized partitions.

Exercise 7-3 Spectral Clustering

(a) Given the dataset from Exercise 7-1, apply spectral clustering to the first ten points (i.e. A - J). When constructing the graph, make sure that each point is connected to its neighbours in an eps = 2 neighbourhood while still having at least two outgoing edges.

(b) As shown in the lecture, spectral clustering uses the Laplacian matrix to determine its clusters. Given an arbitrary graph G and the Laplacian L for G, show that finding an indicator vector f_C that minimizes $\mathrm{f}L\mathrm{f}^T$ leads to an optimal cluster C in G, where

$$\mathrm{f}_C{}^{(i)} = \begin{cases} 1 & \text{if } v_i \in C \\ 0 & \text{else} \end{cases}$$