Dr. Eick

Group Homework Credit Task Group H

**DBSCAN**

*To present on Thursday, October 26, 2023*

**Group H Task DBSCAN**

A dataset consisting of object A, B, C, D, E, F, G, H with the following distance matrix is given:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **A** | **B** | **C** | **D** | **E** | **F** | **G** | **H** |
| **A** |  |  **2** | **7** | **8** | **9** | **10** | **7** |  **4** |
| **B** |  |  | **6** | **7** | **8** | **9** | **8** |  **7** |
| **C** |  |  |  |  **4** | **8** | **9** | **7** |  **8** |
| **D** |  |  |  |  | **9** | **11** | **11** |  **9** |
| **E** |  |  |  |  |  | **3** | **2** |  **10** |
| **F** |  |  |  |  |  |  | **4** |  **11** |
| **G** |  |  |  |  |  |  |  |  **12** |
| **H** |  |  |  |  |  |  |  |  |

Assume DBSCAN is run for this dataset with MINPOINTS[[1]](#footnote-1)=3 and epsilon=ε=5

How many clusters will DBSCAN return and how do they look like? Which objects are core points which objects are outliers/noise points and which objects are border points?

How does the result change if we increase the parameter MINPOINTS from 3 to 4?

Group Homework Credit Task Group I

 **Hierarchical Clustering**

*To present on Thursday, October 26*

**Hierarchical Clustering**

 A dataset consisting of object A, B, C, D, E and F with the following distance matrix is given:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| distance | A | B | C | D | E | F |
| A | 0 | 9 | 8 | 15 | 2 | 11 |
| B |  | 0 | 1 | 6 | 5 | 12 |
| C |  |  | 0 | 7 | 10 | 4 |
| D |  |  |  | 0 | 3 | 13 |
| E |  |  |  |  | 0 | 14 |
| F |  |  |  |  |  | 0 |

a) Assume single[[2]](#footnote-2) link hierarchical clustering is applied to the dataset! What dendrogram will be returned?

b) Does the clustering result change of we use Max/Complete Link[[3]](#footnote-3) instead?

c) How does hierarchical clustering differ from more classical clustering algorithms, such as K-Means and DBSCAN?

1. The object itself counts towards the number of objects in its ε-radius when determining core points! [↑](#footnote-ref-1)
2. When assessing the distance between clusters the minimum distance is used. [↑](#footnote-ref-2)
3. When assessing the distance between clusters the maximum distance is used; that is, the distance of the pair of objects one in one cluster and one in the other cluster that are furthest apart is used. [↑](#footnote-ref-3)