Dr. Eick

**COSC 3337 Fall 2022**

**Group O Group Homework Credit**

**Distance-Based Outlier Detection**

Assume we have a dataset with 10 objects and the following distance matrix

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **A** | **B** | **C** | **D** | **E** | **F** | **G** | **H** | **I** | **J** |
| **A** | **0** | **17** | **19** | **20** | **23** | **25** | **22** | **28** | **23** | **24** |
| **B** |  | **0** | **4** | **5**  | **6** | **7** | **6** | **5** | **29** | **12** |
| **C** |  |  | **0** | **3** | **2** | **4** | **5** | **6** | **33** | **11** |
| **D** |  |  |  | **0** | **4** | **3** | **4** | **5** | **30** | **11** |
| **E** |  |  |  |  | **0**  | **5** | **6** | **7** | **27** | **10** |
| **F** |  |  |  |  |  | **0** | **1** | **2** | **40** | **4** |
| **G** |  |  |  |  |  |  | **0** | **3** | **32** | **14** |
| **H** |  |  |  |  |  |  |  | **0** | **29** | **16** |
| **I** |  |  |  |  |  |  |  |  | **0** | **8** |
| **J** |  |  |  |  |  |  |  |  |  | **0** |

Remark: According to the above distance matrix the distance between object A and object J is 24 and the distance between object C and E is 2!

Describe how a distance-based outlier detection methods works, in general. Next, apply the approach to the above 10-object dataset. Describe the specific results your chosen approach would obtain. Finally, discuss what role hyperparameter(s) play, with respect to the obtained result.