Chair of Computer Science I
Algorithms and Complexity
University of Würzburg

Würzburg, June 22, 2023

Prof. Dr. Alexander Wolff
M.Sc. Marie Diana Sieper

Homework Assignment #6
Algorithms for Geographic Information Systems (SS 2023)

Exercise 1 – Properties of DBSCAN

Consider the definitions underlying DBSCAN for points in the Euclidean plane.

a) Find a finite point set P with two special points p and q, a real $\varepsilon > 0$, and an integer $k \geq 3$ such that p is directly density-reachable from q, but not the other way around. **[2 points]**

b) Find a finite point set P, a real $\varepsilon > 0$, and an integer $k \geq 3$ such that there is a point b in the DBSCAN clustering of P that is a border point for more than one cluster. **[2 points]**

In the following, let X be a finite set of points, let $d\colon X^2 \to \mathbb{R}_{\geq 0}$ be a metric (not necessarily the eucledean metric), let $\varepsilon > 0$, and let $k \leq |X|$. Prove the following claims or give counterexamples.

c) For fixed X, d, and k,

1. there exists a value of ε such that DBSCAN returns no clusters (everything is noise).
2. there exists a value of ε such that DBSCAN returns a single cluster containing all points.

[2 points]

d) For fixed X, d, k, and increasing ε, the number of clusters returned by DBSCAN is nondecreasing. **[1 point]**

e) For fixed X, d, k, and increasing ε, the number of clusters returned by DBSCAN is nonincreasing. **[1 point]**

f) For fixed X, d, $k = 1$, and increasing ε, the number of clusters returned by DBSCAN is nonincreasing. **[1 point]**

g) If $p \in X$ is a border point, there is at least one core point in $N_\varepsilon(p)$. **[1 point]**

Exercise 2 – Faster DBSCAN

Give a linear-time algorithm to compute (the edge set of) $\mathcal{G}_{\text{box}}$. The input consists of the list of strips sorted from left to right, and for each strip, the list of its boxes sorted from bottom to top. Argue correctness and the runtime bound. You may use that the maximum degree in $\mathcal{G}_{\text{box}}$ is at most 22. **[6 points]**

Exercise 3 – DBSCAN in higher dimensions

Let $d > 2$ be a constant, let X be a finite pointset in $\mathbb{R}^d$, let $\varepsilon > 0$, and let $k \in \mathbb{N}$. The problem is to find a set of k clusters in X. How can DBSCAN be adapted to find a solution for this problem? What is the running time of the adaption? **[4 points]**

Please hand in your solutions on WueCampus until 12:00 on Thuesday, June 29.