Prof. Dr. Johannes Blömer

07.01.2016
submission due: 14.01.2016, F1.110, 11:15

Clustering Algorithms

WS 2015/2016

Handout 8

Exercise 1:
Given $D = D_{l_2}$, describe an instance P for which the GonzalesAlgorithm computes a solution C such that

$$\text{cost}_{diam}(C) = (2 - \epsilon) \cdot \text{opt}^{k}_{diam}(P)$$

for any $\epsilon > 0$.

Exercise 2:
Prove that the GonzalesAlgorithm is a 2-approximation algorithm for the radius k-clustering problem.

Exercise 3:
Let C, C' be two distinct density-based clusters of point set P wrt. to ϵ and $MinPts$. Prove that the intersection $C \cap C'$ of C, C' contains only border points.

Exercise 4:
Consider the following version of the DBSCAN algorithm.

$C := \{p \in P \mid |N_\epsilon(p)| \geq MinPts\}$
$N := \{p \in P \mid |N_\epsilon(p)| < MinPts \wedge \nexists c \in C : \; D(p, c) \leq \epsilon\}$
$B := P \setminus (C \cup N)$
Construct a complete graph G with nodes corresponding to C and weight each edge $(c_1, c_2) \in C \times C$ by $D(c_1, c_2)$
Remove all edges of G with weight $> \epsilon$
Compute the connected components of G
for each connected component (denote its set of nodes by C_i) **do**
    Compute all distances between nodes in C_i and in B
    $C_i = C_i \cup \{v \in B | \; \exists c \in C_i : \; D(v, c) \leq \epsilon\}$
return Clustering $C_1, \ldots, C_k$ and N as set of noise points

---

Show that this algorithm computes a valid density-based clustering.