

Problem Set 10

Please hand in your solutions for this problem set via email (roesner@cs.uni-bonn.de) or personally after the lecture until *Tuesday, 8th of January*.

Problem 1

Give an example for a point set $P \subset \mathbb{R}^d$ and a center set $C \subset \mathbb{R}^d$ with $|C| = k$ where its projection $P^{(k)}$ onto a best fit subspace V_k has a very different k -means cost than P , more precisely, find P and C such that

$$\text{dist}^2(P^{(k)}, C) < \frac{1}{2} \text{dist}^2(P, C).$$

Problem 2

Construct an oracle that provides 2-approximate answers to distance queries. More precisely, we want to store $\mathcal{O}(kd)$ information and give an algorithm that based on our information can compute for any C a value $d(C)$ that satisfies

$$\frac{1}{2} \cdot \text{dist}^2(P, C) \leq d(C) \leq 2 \cdot \text{dist}^2(P, C).$$

Problem 3

Consider the following task. You are given d numbers a_1, \dots, a_d which are sorted decreasingly, i.e., $a_1 \geq a_2 \geq \dots, a_d$. You are allowed to store $\text{poly}(k, \epsilon)$ of these numbers plus an extra number N . Then a query comes, and it consists of a set I of k indices. The task is to compute an approximation of the sum of the *other* numbers, i.e., to estimate

$$\sum_{i \in [d], i \notin I} a_i.$$

Design an oracle storing $\text{poly}(k, \epsilon)$ input numbers plus an extra number that is able to guarantee that for every query I , the computed estimate Δ satisfies

$$\left| \Delta - \sum_{i \in [d], i \notin I} a_i \right| \leq \epsilon \cdot \sum_{i \in [d], i \notin I} a_i.$$