Prof. Dr. Johannes Blömer

27.11.2015

submission due: 10.12.2015, F1.110, 11:15

Clustering Algorithms

WS 2015/2016

Handout 5

Exercise 1:

Denote the center of a set $A \subset \mathbb{R}$ by $\mu(A) = \frac{1}{|A|}\sum_{a \in A} a$, and let $opt_1(A)$ be the optimal 1-means cost with respect to A.

Given a set $P \subset \mathbb{R}$, we draw n points uniformly at random from P. Denote by x_i the i-th point that is drawn uniformly at random from P, and let $X = \{x_1, \ldots, x_n\}$. Show that

(a) $E[\mu(X)] = E[x_i] = \mu(P)$

(b) $\mathrm{Var}(\mu(X)) = \frac{1}{n}\mathrm{Var}(x_i)$

(c) With probability $1 - \delta$,

$$|\mu(P) - \mu(X)|^2 < \frac{\mathrm{Var}(x_i)}{n \cdot \delta}.$$

(d) $\mathrm{Var}(x_i) = \frac{1}{|P|} \cdot \mathrm{opt}_1(P)$

(e) With probability $1 - \delta$,

$$cost(P, \mu(X)) < \left(1 + \frac{1}{\delta \cdot n}\right) \mathrm{opt}_1(P).$$

Exercise 2:

Given a set of $P \subset M$ and $k \in \mathbb{N}$ ($|P| \geq k$), we define the discrete k-median problem as follows. Find a subset $C \subseteq P$, $|C| = k$, such that $cost(P, C) = \sum_{p \in P} \min_{c \in C} D_{l_2^2}(c, p)$ is minimized. Denote the optimal discrete k-means cost by $\mathrm{opt}_k^{discr}(P)$.

Let $\mathrm{opt}_k(P)$ be the optimal k-means cost of P. Prove that

$$\mathrm{opt}_k^{discr}(P) \leq 2 \cdot \mathrm{opt}_k(P).$$