Thomas Kesselheim
Alexander Braun

30. Juni 2022
Abgabe: 06. Juli 2022, 9:00 Uhr

Algorithmische Grundlagen des Maschinellen Lernens

Sommersemester 2022

Übungsblatt 10

Aufgabe 1: (5 Punkte)
Sei $S = \{x_1, \dots, x_m\} \subseteq \mathbb{R}^d$. Zeigen Sie, dass zu der optimalen Lösung $C^* = \{c_1^*, \dots, c_k^*\}$ des k-Means-Problems Zentren $C = \{c_1, \dots, c_k\} \subseteq S$ existieren, sodass gilt

$$\phi(S, C) = \sum_{x \in S} \min_{c \in C} \|x - c\|^2 \leq 2 \sum_{x \in S} \min_{c^* \in C^*} \|x - c^*\|^2 = 2\phi(S, C^*).$$

Hinweis: Sie dürfen ohne Beweis ausnutzen, dass für jede Menge $M \subseteq S$, $c \in \mathbb{R}^d$ und $\mu(M) = \frac{1}{|M|} \sum_{x \in M} x$ gilt

$$\phi(M, c) = \sum_{x \in M} \|x - c\|^2 = \sum_{x \in M} \|x - \mu(M)\|^2 + |M| \|\mu(M) - c\|^2.$$

Aufgabe 2: (6 Punkte)
Gegeben eine endliche Menge $S \subset \mathbb{R}^d$ ein Zentrum $c \in \mathbb{R}^d$ und $\alpha \in [0, 1)$. Wir wählen z uniform zufällig aus S. Zeigen Sie, dass mit Wahrscheinlichkeit mindestens α gilt

$$\|z - c\|^2 \leq \frac{1}{1 - \alpha} \cdot \frac{\phi(S, c)}{|S|}.$$

Aufgabe 3: (6 Punkte)
Seien $A, B \subset \mathbb{R}^d$ zwei endliche Mengen mit $A \cap B = \emptyset$. Zeigen Sie, dass

$$\phi(A \cup B, \mu(A \cup B)) = \phi(A, \mu(A)) + \phi(B, \mu(B)) + \frac{|A||B|}{|A| + |B|} \|\mu(A) - \mu(B)\|^2$$

gilt.
Tipp: Wenden Sie Lemma 20.2 jeweils für A und B an.

Aufgabe 4: (3 Punkte)
Zeigen Sie, dass die randomisierte Wahl des ersten Zentrums bei den Algorithmen für k-Means notwendig ist. Konstruieren Sie dazu eine Instanz, für die eine willkürliche (worst-case) Wahl des ersten Zentrums direkt einen beliebig schlechten Zielfunktionswert bezüglich k-Means Zielfunktion liefert, unabhängig davon, wie weitere Zentren gewählt werden.