

5. Übung zur Vorlesung “Bioinformatische Methoden in der Genomforschung”

Sebastian Böcker, Martin Engler

Ausgabe: 26.11.2015

Abgabe: 10.12.2015

Aufgabe 1 (5 Punkte)

Gegeben seien acht Punkte $p_1 = (2, 10)$, $p_2 = (2, 5)$, $p_3 = (8, 4)$, $p_4 = (5, 8)$, $p_5 = (7, 5)$, $p_6 = (6, 4)$, $p_7 = (1, 2)$, $p_8 = (4, 9)$ in der Ebene. Wenden Sie den k-means Algorithmus an, um diese acht Punkte drei Clustern C_1 , C_2 , C_3 zuzuordnen. Initialisieren Sie den Algorithmus mit $C_1 = \{p_1\}$, $C_2 = \{p_2, p_7\}$, $C_3 = \{p_3, p_4, p_5, p_6, p_8\}$. Benutzen Sie dabei die Manhattan-Distanz.

Aufgabe 2 (5 Punkte)

Ein *stationärer Zustand* ist ein Zustand der berechneten Clustern, aus dem der k-means Algorithmus keinen neuen Zustand erreicht. Konstruieren Sie ein möglichst einfaches Beispiel (euklidische Distanz), in dem es mehr als einen stationären Zustand gibt. Konstruieren Sie ein Beispiel mit n Punkten in der Ebene, die k Clustern zugeordnet werden sollen und für die es mindestens 2^k stationäre Zustände gibt.

Aufgabe 3 (5 Punkte)

Beweisen Sie: Ein Graph $G = (V, E)$ ist eine disjunkte Vereinigung von Cliques genau dann, wenn es keine paarweise verschiedenen $u, v, w \in V$ gibt mit $uv \in E$, $uw \in E$, aber $vw \notin E$.

Aufgabe 4 (5 Punkte)

Gegeben sei ein zusammenhängender Graph $G = (V, E)$. Für jedes Paar $uv \in \binom{V}{2}$ seien $s(uv)$ die Kosten für das Hinzufügen der Kante uv , wenn $uv \notin E$, bzw. die Kosten für das Löschen der Kante uv wenn $uv \in E$.

Sei $u, v \in V$ mit $uv \notin E$:

$$s(uv) > \sum_{w \in N(u)} s(uw),$$

wobei $N(u)$ die Menge der benachbarten Knoten von u bezeichnet. Zeigen Sie, dass uv keine Kante der optimalen Lösung des CLUSTER EDITING Problems auf G ist.