5. Exercise "Bioinformatische Methoden in der Genomforschung"

Sebastian Böcker, Martin Hoffmann

Assigned: 23.11.2021
Due: 29.11.2021

Exercise 1 (5 Points)

1. Given eight points on a plane: $p_1 = (2, 10)$, $p_2 = (2, 5)$, $p_3 = (8, 4)$, $p_4 = (5, 8)$, $p_5 = (7, 5)$, $p_6 = (6, 4)$, $p_7 = (1, 2)$, $p_8 = (4, 9)$. Use Lloyd's algorithm to assign these eight points to three cluster C_1, C_2, C_3 Initialize the algorithm with $C_1 = \{p_1\}$, $C_2 = \{p_2, p_7\}$, $C_3 = \{p_3, p_4, p_5, p_6, p_8\}$. Use the manhattan distance.
2. A *steady state* is a state of calculated cluster, from which the lloyd algorithm cannot reach a new, different state. Construct a simple example (using euclidic distance), in which more than one steady state exists. Construct an example with n points on a plane, that shall be assigned to k cluster, and for which at least 2^k steady states exist.

Exercise 2 (5 Points)

Given a connected graph $G = (V, E)$. For every pair $uv \in \binom{V}{2}$, $s(uv)$ are the costs for adding an edge uv, if $uv \notin E$, or the costs for deleting edge uv wenn $uv \in E$.
Let $u, v \in V$ with $uv \notin E$:

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

where $N(u)$ denotes the set of neighboring vertices to u. Show that edge uv is not part of the optimal solution when solving the Cluster Editing problem on G

Exercise 3 (5 Points)

Explain the general sequence of the k-means clustering procecure in your own words (Don't forget running time complexity as well as general assumptions of the algorithm).