

4. Exercise “Bioinformatische Methoden in der Genomforschung”

Sebastian Böcker, Martin Hoffmann

Ausgabe: 16.11.2021

Abgabe: 22.11.2021

Exercise 1 (5 Points)

Use the distance matrices calculated in exercise 3 of last weeks sheet to construct two hierarchical cluster trees: One using *UPGMA* (*unweighted pair group method with arithmetic mean*) and one using *single linkage* as the Intercluster distance. How do the results differ? Which are “better”?

Exercise 2 (5 Points)

Show that *UPGMA* really is “unweighted”

Exercise 3 (10 Points)

Given a distance matrix D on objects $X = \{1, \dots, n\}$, as well as a hierarchical clustering as a rooted tree $T = (V, E)$. To visualize the hierarchical clustering, the leaves of the tree shall be sorted in such a way that the sum of distances of neighbouring leaves in the tree is minimized.

A *linear arrangement* is a permutation π of the numbers $1, \dots, n$. The linear arrangement π is called *consistent* with T , if the arrangement can be produced by flipping inner vertices of T . We are looking for a linear arrangement π , that is consistent with T and minimizes the target function

$$\sum_{j=1}^{n-1} D(\pi(j), \pi(j+1))$$

Find an algorithm that solves above problem in running time complexity of $O(n^5)$. Bonus: Show that your algorithm is in $O(n^4)$