Exercise Sheet 5 for Algorithms of Bioinformatics (Winter 2025/26)

Hand In: *Until 2025-11-21 18:00, on ILIAS.*

Problem 1

40 points

In this exercise, we consider the generalized notion of alignment scores with *affine gap penalties.*

To simplify notation, we assume that $goal = \min$ (i. e., we think of alignment scores as minimal *distance* between strings) and $p(a, b) \geq 0$ for all symbols $a, b \in \Sigma$. Affine gap costs now mean that a *maximal* block of k consecutive inserts or k consecutive deletes contributes $g_0 + k \cdot g$ to the overall alignment score (instead of $k \cdot g$), where $g, g_0 \geq 0$.

Design an algorithm for computing optimal global alignments with affine gap costs. Running time and space complexity should stay the same (same Θ-class) as for the algorithm from class without affine gaps.

Problem 2

40 + 10 + 40 + 10 points

For a given alignment score S, denote by $sim_S : \Sigma^* \times \Sigma^* \to \mathbb{R}$ the score of an optimal alignment w.r.t. S.

Alignment scores of an optimal alignment with $goal_S = \min$ can be interpreted as a measure of *distance* between two strings. Mathematicians have been reasoning about minimal desirable properties that function should have to match our intuition of "distance". This is what they came up with:

A *metric* on a set X is a function $d : X \times X \to \mathbb{R}$ with the following properties for all $x, y, z \in X$:

(M1) $d(x, y) \geq 0$,

(M2) $d(x, y) = 0$ iff $x = y$,

(M3) $d(x, y) = d(y, x)$,

(M4) $d(x, z) \leq d(x, y) + d(y, z)$.

a) Assume that we have $goal_S = \min$, a positive gap penalty $g > 0$ and a transition matrix p that is a metric on Σ.

   Show that sim_S is a metric on $\Sigma^\star$.

b) Give an example for an alignment score where p is *not* a metric on Σ, but sim_S is still a metric on $\Sigma^\star$.

c) Consider now the generalized alignments scores S with *affine gap costs* $g_0 + k \cdot g$ for a deletion or insertion of k consecutive symbols. Assume again that we have $goal_S = \min$, and a transition matrix p that is a metric on Σ. Moreover, $g_0, g \geq 0$ and $g_0 \cdot g > 0$.

   Show that sim_S is a metric on $\Sigma^\star$.

d) Show that string distance measures induced by *semiglobal* or *local* alignments *never* qualify as metric.