TECHNISCHE UNIVERSITÄT KAISERSLAUTERN

Department of Computer Science
Algorithms & Complexity Group

Issue Date: 12.05.2015
Version: 2015-05-12 15:35

Exercise Sheet 2 for Computational Biology (Part 1), SS 15

Hand In: *Until Tuesday, 26.05.2015, 10:00 am, email to* `r_muelle@cs...`*, hand-in box in stairwell 48-6 or in lecture.*

Problem 5

4 points

In this exercise, we consider the generalized notion of alignment scores with *affine gap penalties* (cf. pages 72f in the German lecture notes).

To simplify notation, we assume that $goal_\delta = \min$ (i. e., we think of alignment scores as minimal *distance* between strings) and $\delta(a, b) \geq 0$ 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 $\rho + k \cdot \sigma$ to the overall alignment score (instead of $k \cdot g$), where $\rho, \sigma \geq 0$.

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 6

4 + 1 + 1 + [4] points

Alignments scores of an optimal alignment with $goal_\delta = \min$ can be interpreted as a measure of *distance* between two strings. Mathematicians have been reasoning about minimal desirable properties that every distance measure should have to still 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 $goal_\delta = \min$, a positive gap penalty $g > 0$ and a transition matrix p that is a metric on Σ.

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

b) Give an alignments score where p is *not* a metric on Σ, but sim_δ is still a metric on $\Sigma^\star$.

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

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

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