
Algorithmische Bioinformatik II

Abgabetermin: Freitag, den 20. Dezember, 12⁰⁰

Tutoraufgabe 1 (Vorbereitung bis zum 19.12.19)

Sei $S = \{s_1, \dots, s_k\} \subseteq \Sigma^*$ und sei $M(i) = E_S(s_i)$. Dabei sei ohne Beschränkung der Allgemeinheit $M := M(1) \leq \dots \leq M(k)$.

Zeige, dass $M(\lfloor \frac{k+1}{2} \rfloor) \leq 3M$.

Hinweis: Die Beziehung $\frac{1}{k} \sum_{i=1}^k M(i) < 2M$ kann hilfreich sein (siehe Beweis von Lemma 6.42).

Das Lemma 6.42 bzw. 6.50 darf allerdings nicht direkt verwendet werden.

Aufgabe (Notenbonus) 2

Beweise oder widerlege: Es gibt eine Menge $S \subseteq \Sigma^*$ und eine Kostenfunktion $w : \bar{\Sigma}^2 \rightarrow \mathbb{R}_+$ mit $w(-, -) = 0$, so dass sich das optimale mehrfache Sequenzen-Alignment bzgl. des SP-Distanzmaßes vom optimalen Konsensus-Alignment für S unterscheidet.

Aufgabe (Notenbonus) 3

Verwende den Algorithmus von Carrillo und Lipman zur Berechnung eines Sequenzen-Alignments zwischen zwei Sequenzen $s = ACAT$ und $t = TAGAT$. Hierzu sind für das Distanzmaß die **Gap-Kosten** von 3 und **Mismatch-Kosten** von 2 zu verwenden. Die **globale obere Schranke** für die Distanz von s und t ist mit 6 vorgegeben.

Hinweis: In der Vorlesung wurde dies für 3 oder mehr Sequenzen erläutert, natürlich funktioniert das Verfahren auch mit nur 2 Sequenzen.

Gib die Präfix-Matrix P und die Suffix-Matrix S sowie die kombinierte **Präfix-/Suffix-Matrix** $P + S$ an und **markiere alle Zellen**, die in den **Heap** aufgenommen wurden. Gib dabei ebenfalls die Berechnung der verwendeten **obere Schranke** im Relevanz-Test für das Sequenzpaar (s, t) an.