
Algorithmische Bioinformatik II

Abgabetermin: Freitag, den 23. Dezember, 9⁰⁰ Uhr in Moodle

Tutoraufgabe 1 (Vorbereitung bis zum 21.12.22)

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

Hausaufgabe 2

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).

Die Aussagen der Lemmata 6.42 und 6.50 selbst dürfen allerdings nicht direkt verwendet werden.

Hausaufgabe 3

Betrachte folgende Sequenzen $s_1 = CAGC$, $s_2 = CGTGTC$, $s_3 = CGCGT$ und $s_4 = CGACGT$. Der optimale Abstand für die paarweise Sequenzen-Alignments ist rechts angegeben. Hierbei gilt $w(a, b) = 1$ und $w(a, a) = 0$ für alle $a \neq b \in \overline{\Sigma}$. Konstruiere für diese Sequenzen ein mehrfaches Sequenzen-Alignment mit Hilfe der Center-Star-Methode aus der Vorlesung.

d	s_1	s_2	s_3	s_4
s_1	0	3	3	3
s_2	3	0	2	3
s_3	3	2	0	1
s_4	3	3	1	0