

Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2021

Prof. Dr. Jens Stoye · Dr. Marília D. V. Braga

<https://gi.cebitec.uni-bielefeld.de/teaching/2021summer/sa>

Übungsblatt 4 vom 13.5.2021

Abgabe am 20.5.2021 bis 12:00 Uhr (mittags)

Aufgabe 1 (Eigenschaften von Alignments)

(4 Punkte)

1. Welche Zeichen dürfen in einer Alignmentsspalte bei gegebenem Alphabet Σ vorkommen? Welche Kombination ist nicht erlaubt?
2. Seien $x \in \Sigma^m, y \in \Sigma^n$, A ein globales Alignment von x und y , und sei e eine Editsequenz, die A entspricht. Dann gilt:

$$\max\{m, n\} \leq |A| = |e| \leq m + n.$$

Konstruiere Extremfälle für die Ungleichheiten.

Aufgabe 2 (Beziehung Edit-Sequenz–Globales Alignment)

(4 Punkte)

Gegeben sei die Sequenz $x = \text{ATGCTAG}$.

1. Gib das Alignment A an, welches der Edit-Sequenz $\mathcal{I}_C \text{CCS}_A \mathcal{DCCC} \mathcal{I}_C$, angewandt auf x entspricht und berechne die Projektion $y = \pi_{\{2\}}(A)$.
2. Kann es weitere globale Alignments der Sequenzen x und y geben? Wenn ja, gib eines an, sowie die korrespondierende Edit-Sequenz. Wenn nein, begründe warum es keine geben kann.

Aufgabe 3 (Beziehung Edit-Sequenz–Globales Alignment)

(4 Punkte)

Gegeben sei das Alignment $A = \begin{pmatrix} \text{P} & \text{L} & \text{A} & - & \text{N} & \text{E} & \text{T} & \text{S} \\ \text{P} & \text{A} & \text{R} & \text{E} & \text{N} & - & \text{T} & \text{S} \end{pmatrix}$

1. Berechne die Projektionen $x = \pi_{\{1\}}(A)$ und $y = \pi_{\{2\}}(A)$.
2. Überführe A in eine Edit-Sequenz E von x nach y .
3. Wie hoch sind die Einheits-Kosten der Edit-Sequenz E ?
4. Betrachte nun die folgenden Edit-Scores: $\mathcal{I} = \mathcal{D} = -\frac{1}{2}$, $\mathcal{S} = 0$ und $\mathcal{C} = 1$. Was ist dann der Score von E ?

Aufgabe 4 (Optimaler Edit-Score)

(6 Punkte)

Die Rekurrenz zur Berechnung der Standard-Edit-Distanz mit Einheitskosten (Theorem 3.5 im Skript) lautet für $1 \leq i \leq |x|, 1 \leq j \leq |y|$:

$$D(i, j) = \min \begin{cases} D(i-1, j-1) + \mathbb{1}_{\{x[i] \neq y[j]\}} \\ D(i-1, j) + 1 \\ D(i, j-1) + 1 \end{cases}$$

Die Rekursionsbasisfälle sind:

$$D(0, 0) = 0, \quad D(i, 0) = i \text{ für } 1 \leq i \leq |x| \quad \text{und} \quad D(0, j) = j \text{ für } 1 \leq j \leq |y|$$

1. Betrachte nun die folgenden Edit-Scores: $\mathcal{I} = \mathcal{D} = -\frac{1}{2}$, $\mathcal{S} = 0$ und $\mathcal{C} = 1$. Gib die Rekurrenz und ihre Basisfälle zur Berechnung des optimalen Edit-Scores von zwei Sequenzen x und y an.
2. Benutze die Rekurrenz zur Berechnung des optimalen Edit-Scores der Sequenzen $x = \text{PLANETS}$ und $y = \text{PARENTS}$ mithilfe einer Edit-Matrix. Benutze außerdem eine weitere Matrix, in der die optimalen Edit-Operationen gespeichert werden, um alle optimalen Edit-Sequenzen bestimmen zu können. Schreibe eine optimale Edit-Sequenz explizit auf und gib das korrespondierende Alignment an.