

# Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2020

Prof. Dr. Jens Stoye · Dr. Marília D. V. Braga · Leonie R. Brockmann · Rebecca K. Pfeil

<https://gi.cebitec.uni-bielefeld.de/teaching/2020summer/sa>

## Übungsblatt 10 vom 25.06.2020

Abgabe am 2.07.2020 bis 12:00 Uhr (mittags)

### Aufgabe 1 (Sum-of-Pairs Multiple Alignment)

(4 Punkte)

Gegeben sei das folgende Alignment  $A$  von 4 Sequenzen:

$$A = \begin{pmatrix} A & T & - & - & - & C & G \\ - & T & - & - & C & C & G \\ A & A & A & - & C & C & - \\ - & T & A & T & C & - & G \end{pmatrix}$$

Betrachte die folgende Score-Funktion: match = 4, mismatch = -1 sowie affine Gapkosten mit gap-open = 2 und gap-extension = 1. Berechne den Sum-of-Pairs-Score von  $A$ . Gib alle Zwischenschritte deiner Berechnung an.

### Aufgabe 2 (Sum-of-Pairs Multiple Alignment)

(6 Punkte)

Gegeben seien die Sequenzen  $s_1 = GCT$ ,  $s_2 = ACT$  und  $s_3 = AGC$ . Betrachte die folgende Score-Funktion: match = 2 und mismatch = gap = -1.

1. Berechne für jedes der drei Paare  $(s_1, s_2)$ ,  $(s_1, s_3)$  und  $(s_2, s_3)$  die zweidimensionale Alignment-Matrix und ein optimales globales Alignment.
2. Finde ein globales multiples Alignment  $A$  von  $s_1$ ,  $s_2$  und  $s_3$ , dessen Sum-of-Pairs-Score  $S_{SP}(A) = 6$  ist.
3. Berechne die drei Projektionen  $\pi_{\{1,2\}}$ ,  $\pi_{\{1,3\}}$  und  $\pi_{\{2,3\}}$  des Alignments  $A$  und ihre Scores. Welche der Projektionen korrespondieren zu dem oben berechneten optimalen Alignment und welche nicht?

### Aufgabe 3 (Suchraum multipler Alignments)

(6 Punkte)

Um den Wert  $D(i_1, i_2, \dots, i_k)$  eines Knotens im  $k$ -dimensionalen Edit-Graphen mittels des grundlegenden Algorithmus (Abschnitt 11.2.1 im Skript) zu berechnen, muss über eine Anzahl an Werten minimiert werden, die der Anzahl Vorgängerknoten im Graphen entspricht. Die Zahl dieser Vorgängerknoten ist unterschiedlich, je nachdem, wo man sich im Edit-Graphen befindet; es gibt also verschiedene *Typen* von Knoten. Betrachte die folgenden Situationen:

- (i)  $k = 2$  Sequenzen der Länge 4;
- (ii)  $k = 3$  Sequenzen der Länge 4; und
- (iii)  $k = 4$  Sequenzen der Länge 4.

Für die entsprechenden  $k$ -dimensionalen Edit-Graphen sind die folgenden Fragen zu beantworten:

1. Wie viele Knoten hat der  $k$ -dimensionale Edit-Graph? (Start- und Endknoten des Graphen können ignoriert werden.)
2. Welche verschiedenen Typen an Knoten gibt es, und wie viele eingehende Kanten hat jeder von ihnen?
3. Wie viele Knoten jeden Typs gibt es?
4. Wie viele Berechnungsschritte werden also insgesamt durchgeführt?

**Aufgabe 4 (Carrillo-Lipman-Heuristik)**

(4 Punkte)

1. Wie viele Carrillo-Lipman-Schranken  $U_{x,y}$  müssen berechnet werden, um  $k$  Sequenzen zu alignieren? Wie ist die asymptotische Laufzeit für die Berechnung der Schranken?
2. Für welche Sequenzen funktioniert die Carrillo-Lipman-Heuristik gut bzw. schlecht?

**Aufgabe 5 (NP-vollständige Probleme)**

(2\* Punkte)

Gegeben sei ein NP-vollständiges Problem  $Q$ . Was kann über die Komplexität eines anderen Problems  $Q'$  gesagt werden, wenn ...

1. ...  $Q' \leq_p Q$ ?
2. ...  $Q \leq_p Q'$ ?

**Aufgabe 6 (Umgang mit NP-vollständigen Problemen)**

(6\* Punkte)

Angenommen, ein Problem ist NP-vollständig.

1. Wie kannst du vorgehen, um zu einer Lösung zu kommen? Zähle fünf Möglichkeiten auf, die du je mit einem kurzen Satz beschreibst.
2. In welchen Fällen ist dein Ergebnis noch korrekt? Was kannst du über eventuelle Abweichungen vom korrekten Ergebnis sagen?
3. Wenn du in der Lage wärst zu zeigen, dass sich dein NP-vollständiges Problem in polynomieller Zeit lösen lässt, welche Folge hätte das für andere NP-vollständige Probleme?