Ludwig-Maximilians-Universität München
Institut für Informatik
Prof. Dr. Volker Heun

Wintersemester 2019/20
Semestralklausur
13. Februar 2020

Algorithmische Bioinformatik II

Vorname	Name	Matrikelnummer

Reihe	Platz	Unterschrift

Hiermit stimme ich einer Veröffentlichung meines Klausurergebnisses dieser Semestralklausur unter Verwendung meiner Matrikelnummer im Internet zu. *Ja* ☐ *Nein* ☐

____
(Unterschrift)

Allgemeine Hinweise zur Semestralklausur

- Vor der Prüfung ist diese Seite mit Vornamen, Namen, Matrikelnummer, Reihe und Platz **leserlich mit Druckbuchstaben** zu versehen und zu unterschreiben.
- Bitte **nicht** in *roter* oder *grüner* Farbe bzw. **nicht** mit Bleistift schreiben.
- Der Studentenausweis und ein amtlicher Lichtbildausweis sind bereit zu halten.
- Die reine Bearbeitungszeit beträgt **120 Minuten**.
- Es sind insgesamt 40 Punkte zu erreichen, zum Bestehen sind 17 Punkte nötig.

Viel Erfolg!

Hörsaal verlassen von .... bis .... von .... bis ....

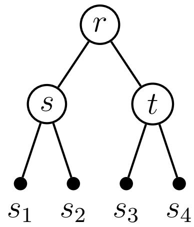
Vorzeitig abgegeben um ....

	Hz	A1	A2	A3	A4	A5	$\sum$
Erstkorrektur							
Nachkorrektur							
Zweitprüfer							

Vorname: ______________ Name: ______________ Matrikelnummer: ____________

Aufgabe 1 (8 Punkte)

Berechne für den rechts angegebenen vollständigen Baum ein optimales **uniform** geliftetes Alignment mittels der dynamischen Programmierung.



d	s_1	s_2	s_3	s_4
s_1	0	1	5	6
s_2		0	2	7
s_3			0	3
s_4				0

Vorname: ______________ Name: ______________ Matrikelnummer: ____________

Aufgabe 2 (8 Punkte)

Verwende den Algorithmus von Carrillo und Lipman zur Berechnung eines Sequenzen-Alignments zwischen zwei Sequenzen $s = TATA$ und $t = ATG$. 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 9 vorgegeben.

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

Gib die kombinierte **Prefix-/Suffix-Matrix** $P + S$ und dessen Herleitung 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.

Vorname: ______________ Name: ______________ Matrikelnummer: ____________

Aufgabe 3 (8 Punkte)

Bestimme für die folgenden Blöcke von Sequenzen die zugehörigen Häufigkeiten $H(a, b)$ für die BLOSUM50-Matrix.

$$
\begin{array}{ll}
s_1^{(1)} = \text{CABCC} & s_1^{(2)} = \text{CBBCACB} \\
s_2^{(1)} = \text{BAACB} & s_2^{(2)} = \text{CCBCABC} \\
s_3^{(1)} = \text{CCABB} & s_3^{(2)} = \text{BCBBABB} \\
s_4^{(1)} = \text{CAACB} & s_4^{(2)} = \text{ABAACBB}
\end{array}
$$

Vorname: ______________ Name: ______________ Matrikelnummer: ____________

Aufgabe 4 (8 Punkte)

Wir betrachten eine Münze, wobei mit Wahrscheinlichkeit $p \in (0, 1]$ Kopf erscheint und und mit Wahrscheinlichkeit $1 - p$ Zahl. Sei X die Zufallsvariable, die zählt, wie oft die Münze geworfen werden muss bis Kopf erscheint, dann gilt

$$\mathrm{Ws}[X = N] = p \cdot (1 - p)^{N-1}.$$

a) Gib die allgemeinen Formeln sowohl für den Maximum-Likelihood-Schätzer als auch den Maximum-A-Posteriori-Schätzer an.

b) Angenommen die Münze wurde N-mal geworfen, bis das erste Mal Kopf erschien. Bestimme die Likelihood-Funktion für dieses Ergebnis und gib dann den Maximum-Likelihood-Schätzer für p an.

c) Angenommen die Münze wurde N-mal geworfen, bis das erste Mal Kopf erschien. Bestimme die Posteriori-Wahrscheinlichkeit für dieses Ergebnis bezüglich des Parameterraums $p \in (0, 1]$, wobei der Prior $f(p) = 2p$ ist und gib dann den Maximum-A-Posteriori-Schätzer für p an.

Vorname: ______________ Name: ______________ Matrikelnummer: ____________

Aufgabe 5 (8 Punkte)

Max3Cut

Eingabe: Ein ungerichteter Graph $G = (V, E)$

Lösung: Ein Partition V_1, V_2, V_3 von V, d.h. $V_1 \cup V_2 \cup V_3 = V$ und $V_i \cap V_j = \emptyset$ für alle $i \neq j \in [1 : 3]$

Optimum: Maximiere $\sum_{i=1}^{3} \sum_{j=i+1}^{3} |(V_i \times V_j) \cap E|$.

Hierbei ist $(V_i \times V_j) = \{\{v, w\} : v \in V_i \wedge w \in V_j\}$.

Anschaulich ist die Anzahl der Kanten zu maximieren, die zwischen den Mengen der Partition der Knoten verlaufen.

a) Zeige, dass Max3Cut $\in \mathcal{NPO}$.

b) Konstruiere eine polynomielle 3-Approximation für Max3Cut.

Hinweis: Korrektheitsbeweise und Laufzeitanalyse nicht vergessen.