Ludwig-Maximilians-Universität München | Wintersemester 2022/23
Institut für Informatik | Wiederholungsklausur
Prof. Dr. Volker Heun | 20. April 2023

Algorithmische Bioinformatik II

Vorname	Name	Matrikelnummer
Reihe	**Platz**	**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)

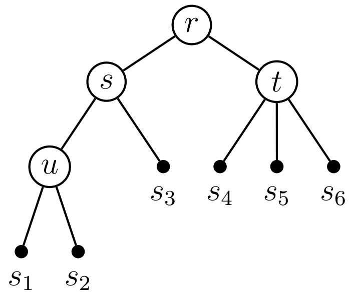
Verwende den Algorithmus von Carrillo und Lipman zur Berechnung eines Sequenzen-Alignments zwischen zwei Sequenzen $s = TAGA$ und $t = CAA$. Hierzu sind für das Distanzmaß die **Gap-Kosten** von 3 und **Mismatch-Kosten** von 2 zu verwenden. Die **globale obere Schranke** für die Distanz von s und t ist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

Hinweis: In der Vorlesung wurde dies für 3 oder mehr Sequenzen besprochen, 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 2 (8 Punkte)



d	s_1	s_2	s_3	s_4	s_5	s_6
s_1	0	1	3	2	2	3
s_2		0	2	2	3	3
s_3			0	3	3	3
s_4				0	1	1
s_5					0	2
s_6						0

Berechne für den rechts angegebenen Baum und der zugehörigen Distanzmatrix **ein** optimales geliftetes phylogenetisches mehrfaches Sequenzen-Alignment gemäß der dynamischen Programmierung. Hierzu müssen nur die gelifteten Sequenzen für jeden inneren Knoten angegeben werden (nicht das mehrfache Sequenzen-Alignment selbst). Es müssen nur legale Paare berücksichtigt werden.

Vorname: ______________ Name: ______________ Matrikelnummer: ____________

Aufgabe 3 (8 Punkte)

Sei X die Zufallsvariable, die zählt, wie oft man eine Münze werfen muss, bis zum ersten Mal Kopf erscheint. Für $k \in \mathbb{N}$ gilt $\mathrm{Ws}[X = k] = (1-p)^{k-1}p$, wenn p die Wahrscheinlichkeit ist, dass bei einem Wurf Kopf erscheint.

Angenommen, bei einem Versuch wird 6-mal geworfen bis zum ersten Mal Kopf erscheint. Überprüfe mit Hilfe des Likelihood-Ratio-Tests die Null-Hypothese ($p = \frac{1}{2}$) gegen die Alternativ-Hypothese ($p = \frac{1}{4}$) für das Signifikanz-Niveau $\alpha = 0.02$.

Hilfe: Für $F(r,p) = \sum_{i=r}^{\infty}(1-p)^{i-1}p = (1-p)^{r-1}$ gilt:

r	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$F(r,1/4)$	1.00	0.75	0.56	0.42	0.32	0.24	0.18	0.13	0.10	0.08	0.06	0.04	0.03	0.02	0.02	0.01
$F(r,1/2)$	1.00	0.50	0.25	0.12	0.06	0.03	0.02	0.01	0.004							
$F(r,3/4)$	1.00	0.25	0.06	0.02	0.004											

Vorname: ______________ Name: ______________ Matrikelnummer: ____________

Aufgabe 4 (8 Punkte)

Sei $Z_n \in [0 : 1]$ für $n \in \mathbb{N}_0$ eine Zufallsvariable, die den Ausgang des n-ten Wurfs einer Münze beschreibt (wobei beide Ausgänge gleichwahrscheinlich sind).

Betrachte die Markov-Kette $X_0 := Z_0$ und für $n > 0$

$$X_n := Z_n + 2 \cdot Z_{n-1}.$$

a) Bestimme das zu $(X_n)_{n\in\mathbb{N}}$ gehörige Markov-Modell (Q, P, π).

b) Zeige, dass das Markov-Modell aus a) ergodisch ist.

c) Gib die stationäre Verteilung des Markov-Modells aus a) an.

Vorname: ________________ Name: ________________ Matrikelnummer: ______________

Aufgabe 5 (8 Punkte)

Betrachte das unten angegebene MAX2SAT-Problem.

a) Zeige, dass MAX2SAT $\in \mathcal{NPO}$.

b) Konstruiere eine polynomielle 2-Approximation für MAX2SAT.

Beachte: Das zu MAX2SAT gehörige Entscheidungsproblem ist $\mathcal{NP}$-hart.

Hinweise: Versuche einen Greedy-Algorithmus über die Variablen, die in F vorkommen, zu entwerfen. Korrektheitsbeweise und Laufzeitanalyse nicht vergessen.

MAX2SAT

Eingabe:	Ein Boolesche Formel $F = \bigwedge_{i=1}^{k} C_i$ in konjunktiver Normalform, wobei jede Klausel aus genau 2 verschiedenen Literalen besteht.
Lösung:	Ein Belegung $B : V(F) \to \mathbb{B}$
Optimum:	Maximiere $\mid \{i \in [1:k] \, : \, B(C_i) = 1\}$.