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

Wintersemester 2016/17
Semestralklausur
16. Februar 2017

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)

Gegeben seien $s_1 = \text{TACGG}$, $s_2 = \text{ACTGG}$, $s_3 = \text{ATGG}$ und $s_4 = \text{ATGGC}$. Konstruiere für diese Sequenzen ein mehrfaches Alignment mit Hilfe der Center-Star-Methode.

Hierbei gilt $w(a, b) = 1$ und $w(a, a) = 0$ für alle $a \neq b \in \overline{\Sigma}$.

Lücken sollen dabei wiederverwendet werden.

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

Vorname: ________________ Name: ________________ Matrikelnummer: ______________

Aufgabe 2 (8 Punkte)

Betrachte die Sequenzen $s_1 = \text{CAT}$, $s_2 = \text{ATA}$ und $s_3 = \text{TCA}$. Berechne die C-optimalen Schnittpositionen mit Respekt zu $c_1 = 1$ und die daraus resultierenden mehrfachen Alignments gemäß des Divide-and-Conquer-Alignment-Algorithmus, wobei nach der **ersten** Rekursion bereits jeweils ein optimales Alignment für die jeweiligen Präfixe bzw. Suffixe berechnet wird.

Für die Kostenfunktion des SP-Distanzmaßes gelte $w(a, a) = 0$ und $w(a, b) = 1$ für alle $a \neq b \in \overline{\Sigma}$.

Vorname: ______________ Name: ______________ Matrikelnummer: ____________

Aufgabe 3 (8 Punkte)

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

$$
\begin{array}{ll}
s_1^{(1)} = \mathrm{ACCCA} & s_1^{(2)} = \mathrm{CBBCACBAC} \\
s_2^{(1)} = \mathrm{ABACA} & s_2^{(2)} = \mathrm{CCBCABCAA} \\
s_3^{(1)} = \mathrm{CBABA} & s_3^{(2)} = \mathrm{BCBBABBAB} \\
s_4^{(1)} = \mathrm{ACABB} & s_4^{(2)} = \mathrm{CBACABBBA}
\end{array}
$$

Vorname: ______________ Name: ______________ Matrikelnummer: ____________

Aufgabe 4 (8 Punkte)

Sei $Z_n \in \{0, 1\}$ eine Zufallsvariable, die den Ausgang des n-ten Wurfs einer Münze beschreibt (wobei die beiden Ausgänge gleichwahrscheinlich sind).

Betrachte $X_0 := 2 \cdot Z_0$ und für $n > 0$

$$X_n := 3 \cdot Z_n + 2 \cdot Z_{n-1}.$$

a) Begründe, dass $(X_n)_{n\in\mathbb{N}}$ eine Markov-Kette erster Ordnung ist.

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

c) Gib für das Markov-Modell aus b) die stationäre Verteilung an.

d) Begründe, dass das Markov-Modell aus b) ergodisch ist

Vorname: ______________ Name: ______________ Matrikelnummer: ____________

Aufgabe 5 (8 Punkte)

Zeige, dass MinEdgeCover $\leq_{\text{PTAS}}$ MinSAT. Gib dazu explizit eine PTAS-Reduktion (f, g, α) an und weise die erforderlichen Eigenschaften einer PTAS-Reduktion nach.

MinEdgeCover (MinEC)

Eingabe: Ein ungerichteter Graph $G = (V, E)$ mit $E = \{e_1, \ldots, e_m\}$.
Lösung: Eine Teilmenge $E' \subseteq E$ der Kanten, so dass jeder Knoten mindestens ein Endpunkt einer Kante in E' ist, d.h. $\forall v \in V : \exists e \in E' : e \cap \{v\} \neq \emptyset$.
Optimum: Minimiere $|E'|$.

MinSAT

Eingabe: Eine Boolesche Formel F über $V(F) = X$.
Lösung: Eine erfüllende Belegung $B : X \to \mathbb{B}$, d.h. $B(F) = 1$.
Optimum: Minimiere $\mu(B) = |\{x \in X : B(x) = 1\}|$.

Achtung: Bei MinSAT ist **nicht** die Anzahl erfüllter Klauseln zu minimieren, sondern die Anzahl der auf wahr gesetzten Variablen.