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

Wintersemester 2016/17
Wiederholungsklausur
03. Mai 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 ....

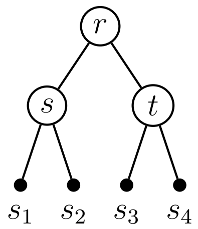
Vorzeitig abgegeben um ....

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

Vorname: ____________ Name: ____________ Matrikelnummer: ____________

Aufgabe 1 (8 Punkte)

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	3	1	3
s_2		0	4	5
s_3			0	2
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 $s = ATTG$ und $t = TG$. 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 7 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)

Betrachte die folgende Matrix mit den Werten $n_{a,b}$ für die PAM-Matrizen.

$n_{a,b}$	A	B	C
A	0	100	200
B	100	0	200
C	200	200	0

a) Gib die allgemeine Formel für 1-PAM-Matrizen für die Werte $w(\cdot,\cdot)$ an:

b) Berechne die Werte $w(\cdot,\cdot)$ der zugehörigen 1-PAM.

c) Was bedeutet PAM? Was ist der prinzipielle Unterschied zwischen einer 2-PAM und einer Substitutionsmatrix basierend auf einer Mutationsrate von 2%?

Hinweis: Gib auch die Zwischenergebnisse an. Am Ende müssen die Logarithmen nicht ausgerechnet werden, auftretende Brüche aber weitestgehend gekürzt werden.

Vorname: ______________ Name: ______________ Matrikelnummer: ____________

Aufgabe 4 (8 Punkte)

Wir betrachten vier verschiedene Typen von Münzen A_i für $i \in \{1,2,3,4\}$. Eine Münze vom Typ A_i zeigt mit Wahrscheinlichkeit $\frac{i}{10}$, Kopf und mit der entsprechenden Gegenwahrscheinlichkeit Zahl. Wir wählen zufällig eine Münze (aus den vier Typen) und erhalten anschließend beim viermaligen Werfen dieser Münze genau zweimal Kopf.

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

b) Bestimme die Likelihood-Funktion für dieses Ergebnis bezüglich des Parameterraums $\Theta = \{A_1, A_2, A_3, A_4\}$ und gib dann den Maximum-Likelihood-Schätzer für den Typ der gewählten Münze aus dem Parameterraum $\Theta = \{A_1, A_2, A_3, A_4\}$ an.

c) Bestimme die Posteriori-Wahrscheinlichkeit für dieses Ergebnis bezüglich des Parameterraums $\Theta = \{A_1, A_2, A_3, A_4\}$ und gib dann den Maximum-A-Posteriori-Schätzer für den Typ der gewählten Münze aus dem Parameterraum $\Theta = \{A_1, A_2, A_3, A_4\}$ an, wobei die Priori-Wahrscheinlichkeit, dass wir einen Würfel vom Typ A_i zufällig ziehen, bei $\frac{5-i}{10}$ liegt.

Vorname: __________ Name: __________ Matrikelnummer: __________

Aufgabe 5 (8 Punkte)

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

MinTSP (Undirected Traveling Salesperson)

Eingabe: Ein vollständiger ungerichteter Graph $G = (V, E)$ und Kantengewichten w.
Lösung: Ein ungerichteter Hamiltonscher Kreis, d.h. eine Permutation $(v_{\pi(1)}, \ldots, v_{\pi(n)})$ der Knotenmenge V mit $\{v_{\pi(i)}, v_{\pi((i \bmod n)+1)}\} \in E$ und $n = |V|$.
Optimum: Minimiere $\sum_{i=1}^{n} w(v_{\pi(i)}, v_{\pi((i \bmod n)+1)})$ mit $n = |V|$.

MinDirectedTSP (Directed Traveling Salesperson)

Eingabe: Ein vollständiger gerichteter Graph $G = (V, E)$ und Kantengewichten w.
Lösung: Ein gerichteter Hamiltonscher Kreis, d.h. eine Permutation $(v_{\pi(1)}, \ldots, v_{\pi(n)})$ der Knotenmenge V mit $(v_{\pi(i)}, v_{\pi((i \bmod n)+1)}) \in E$ und $n = |V|$.
Optimum: Minimiere $\sum_{i=1}^{n} w(v_{\pi(i)}, v_{\pi((i \bmod n)+1)})$ mit $n = |V|$.