

Zweite Klausur zur Übung Algorithmische Bioinformatik

Freie Universität Berlin, WS 2004/05

Utz J. Pape · Ben Rich · Dr. Stefan Röpcke · Prof. Dr. Martin Vingron

Name:

Matrikelnummer:

1. Wieviele binäre gewurzelte Bäume mit $n = 5$ Blättern gibt es? [2 Punkte]

2. Gegeben ist folgende Distanzmatrix für die Taxa A, B und C:

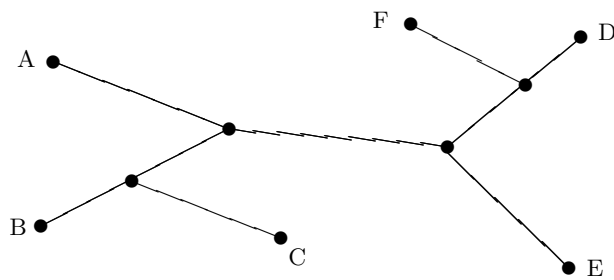
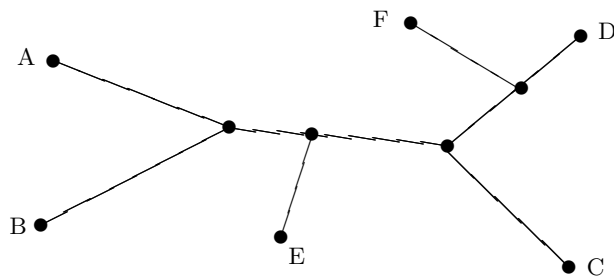
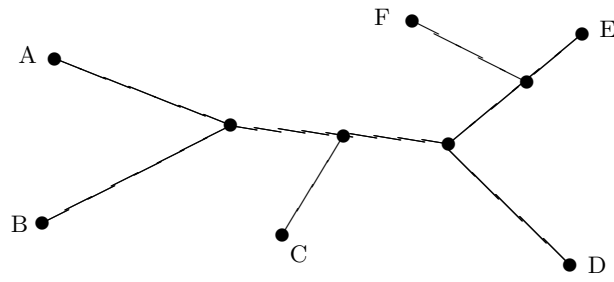
$$\begin{pmatrix} 0 & 7 & 10 \\ 7 & 0 & 11 \\ 10 & 11 & 0 \end{pmatrix}$$

Zeigen Sie, dass diese keine Ultrametrik enthält! [2 Punkte]

3. Wie lautet die 4-Punkte Bedingung und was überprüft man mit ihr? [2 Punkte]

4. Welche Baumtopologie verwendet Neighbor-Joining bei der Initialisierung? [1 Punkt]

5. Bestimmen Sie die nicht-trivialen Splits der drei folgenden Bäume [3 Punkte]



Bestimmen sie den Konsensus Baum nach der majority-rule mit 50%. [1 Punkt]

6. Berechnen Sie mit dem Fitch Algorithmus den maximum parsimony score von dem Baum mit dem Split $ab|c$, wobei die Nukleotide für die Taxa $a=G$, $b=C$ und $c=G$ lauten. [2 Punkte]

7. Gegeben sind folgende Sequenzen von vier den Taxa a , b , c und d :

- (a) ACG
- (b) ACC
- (c) AGC
- (d) ACG

Welche Sites sind für ungewurzelte Bäume uninformativ? [2 Punkte]

8. Wie sieht die Ratenmatrix in der allgemeinen Form zum 2-Parameter-Kimura Modell aus? [2 Punkte]

Was sind die beiden Parameter und welche biologischen Prozesse beschreiben sie? [2 Punkte]

9. In einem Alignment der Länge 30 sind 10 Mismatches enthalten. Schätzen Sie die Distanz der beiden Sequenzen mit dem naiven Schätzer. [1 Punkt]

Ist der Jukes-Cantor korrigierte Wert höher oder niedriger? [1 Punkt]

Wofür korrigiert die Jukes-Cantor Korrektur? [1 Punkt]

10. Mit welchem Algorithmus berechnet man in einem HMM die Wahrscheinlichkeit der Beobachtung O gegeben dem Modell λ ? Geben Sie auch die Formel dazu an! [2 Punkte]

11. Wie lautet der conditional maximum likelihood Schätzer für labelled HMMs (Formel)?
[1 Punkt]

12. Sie haben eine Sequenz gegeben und wollen diese mit einer Markovkette modellieren. Da Sie nicht wissen, ob Sie eine Markovkette 0ter oder 1ter Ordnung nehmen sollten, wollen Sie diese Entscheidung mit einem statistischen Test vornehmen.

Wie lautet die Nullhypothese (in Formeln)? [1 Punkt]

Welcher Verteilung folgt die Pearson-Differenz-Statistik unter der Nullhypothese? [1 Punkt]

13. Sie modellieren eine Längenverteilung mit einer Markovkette mit 3 Zuständen, wobei der Selbstübergang p und der Übergang zum nächsten Zustand $1 - p$ ist. Was für einer Verteilung folgen die Längen der generierten Sequenzen? [1 Punkt]

Wie lautet die entsprechende Formel? [1 Punkt]