Übungen zur Vorlesung Algorithmische Bioinformatik

Freie Universität Berlin, WS 2014/15
Martin Vingron · Juliane Perner · Annkatrin Bressin

Blatt 3 · Ausgabe am 27.10.2014
Abgabe am 03.11.2014 vor Beginn der Vorlesung

Name: Matrikelnummer: Übungsgruppe:

Aufgabe 1 (*20 Punkte; Theorie*)**.** Gegeben sind die folgenden Distanzmatrizen:

(a)

	a	b	c	d
a	0	8	9	5
b		0	9	8
c			0	9
d				0

(b)

	a	b	c	d
a	0	7	2	8
b		0	8	7
c			0	9
d				0

1. Überprüfen Sie ob diese Matrizen eine additive Metrik oder eine Ultrametrik repräsentieren.

2. Wenden Sie UPGMA und Single linkage clustering auf die Matrizen an. Ist das Clustering in allen Fällen eindeutig? Haben Sie eine Erklärung dafür?

Aufgabe 2 (*10 Punkte; Praxis*)**.** Installieren Sie die Software PHYLIP[1] zum Rekonstruieren von phylogenetischen Bäumen. Nutzen Sie PHYLIP einmal mit dem Maximum-Parsimony und einmal mit dem Maximum Likelihood Algorithmus um einen phylogenetischen Baum für die mitochondriale DNA-Sequenzen[2] zu konstruieren. Geben Sie die resultierenden Bäume im Newick-Format an und kommentieren Sie kurz die Ergebnisse.

Aufgabe 3 (*30Punkte; Programmieren*)**.** Erstellen Sie ein Programm welches...

1. ein Alignment im FASTA-Format einliest.

2. aus diesem Alignment und mittels der Jukes-Cantor Korrektur mit normierter Hamming-Distanz eine Distanzmatrix erstellt. Betrachten Sie dabei die Gaps „-“ als gleichwertige Mutation.

3. aus dieser Distanzmatrix mit dem UPGMA-Algorithmus einen phylogenetischen Baum im Newick-Format erzeugt.

4. mit folgendem Befehl aufrufbar ist: program_name inputfile outputfile

Testen Sie ihr Programm anhand des auf der Vorlesungsseite gegebenen Alignmentfiles[3] und zeichnen Sie den resultierenden Baum mit PHYLIP. Dazu benötigen Sie ein fontfile, welches Sie im src-Ordner von PHYLIP finden.

[1] http://evolution.genetics.washington.edu/phylip.html
[2] Material 1: http://www.molgen.mpg.de/Algorithmische-Bioinformatik-WS1415/u3/mitoDNA
[3] Material 2: http://www.molgen.mpg.de/Algorithmische-Bioinformatik-WS1415/u3/test88

Aufgabe 4 (*40Punkte; Theorie*)**.** Wir haben eine Münze zehn mal geworfen und dabei 7mal Kopf und 3mal Zahl beobachtet. Wir möchten nun wissen ob die Münze fair ist. Berechnen Sie dazu den Maximum-Likelihood-Schätzer für einen Münzwurf unter der Annahme, dass unsere Zufallsvariable X (in unserem Fall ist $X = 3$) binomialverteilt ist:

$$f(x;p) = P_p(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

1. Bestimmen Sie die Likelihood-Funktion $L(p) = P(\mathrm{Data}|p)$ für $p \in [0, 1]$.
2. Stellen Sie L für $p \in [0..1]$ grafisch dar. Vergleichen Sie ihren Plot mit der Likelihood-Funktion für den Fall, dass wir 30mal Zahl und 70mal Kopf beobachtet hätten. Was fällt auf?
3. Finden Sie das Maximum der Likelihood Funktion $L(p)$, in dem Sie L nach p ableiten. Bestimmen Sie auch das Maximum der log-Likelihood Funktion $l(p) = \log[L(p)]$. Was fällt auf?
4. Berechnen Sie auch die Wahrscheinlichkeit, dass die Daten beobachtet werden, wenn $p = 0.5$ ist. Erklären Sie anhand dieser, ob die Münze fair ist.