Prof. Dr. Knut Reinert
René Rahn
Jakob Schulze
Kathleen Gallo
Denise Thiel

Institut für Informatik
AG Algorithmische Bioinformatik

Algorithmen und Datenstrukturen in der Bioinformatik
Zwölftes Übungsblatt WS 14/15

Abgabe Donnerstag 12:00 Uhr

Niveau I

Aufgabe 1: Viterbi und posterior decoding

Betrachten Sie das HMM mit den Transitionswahrscheinlichkeiten A und den Emissionswahrscheinlichkeiten e:

A:=

	0	P	Q
0	0	0.2	0.8
P	0.5	0.3	0.2
Q	0.7	0.3	0

e:=

	x	y
P	0.5	0.5
Q	0.2	0.8

a) Berechnen Sie fr die Sequenz *yxy* den 'Viterbi'- und den 'posterior decoding'-Pfad.

b) Erklären Sie, warum sich die beiden unterscheiden. Warum kann das Ergebnis des "falschen"Pfades trotzdem von Interesse sein?

Aufgabe 2: Training von HMM's

Gegeben die Sequenz *cdcccd* und die dazugehörige Zustandsfolge $AABBAB$

a) Bestimmen sie mit der *maximum likelihood*-Methode die Parameter für einen HMM, der diese Sequenz erzeugt

b) Welche Probleme können auftreten, wenn man einen HMM mit zu wenigen Sequenzen trainiert? Wie kann man diese beheben?

Niveau II

Aufgabe 3: CpG-Inseln

'C'- und 'G'-reiche DNA-Regionen werden CpG-Inseln genannt. Allerdings gibt es keine eindeutige Definition. CpG-Inseln sind öfter Methyliert als andere Regionen, haben also mehr Methyl-Gruppen. Überdurchschnittliche Methylierung wird als Hypermethylierung bezeichnet, unterdurchschnittliche als Hypomethylierung. Nun versucht man, über die Methylierungslevel einer Sequenz vorherzusagen, ob sie eine CpG-Insel ist.

Modellieren Sie das Problem als HMM erster Ordnung. Gehen Sie davon aus, dass 10% der DNA (unendlich lang) zu CpG-Inseln gehören und dass man mit einer Wahrscheinlichkeit von 5% von 'C'- und 'G'-armen Regionen in eine CpG-Insel kommt. In CpG-Inseln soll die Wahrscheinlichkeit für normale Methylierung doppelt so hoch sein wie für Hypermethylierung, Hypomethylierungen tauchen hier nicht auf. Außerhalb von CpG-Inseln gibt es keine Hypermethylierung. Zwei Drittel der DNA sind normal methyliert.

a) Geben Sie die Parameter des HMM an.

b) Was ist die Wahrscheinlichkeit, dass ein normal methylierter Abschnitt zu einer CpG-Insel gehört?

c) Was ist die Wahrscheinlichkeit, dass ein normal methylierter Abschnitt zu einer CpG-Insel gehört, wenn die Region davor Hypermethyliert ist?