

# Übungen zur Vorlesung Algorithmische Bioinformatik

Freie Universität Berlin, WS 2008/09  
Roland Krause · Hannes Luz · Utz J. Pape · Martin Vingron

**Blatt 5a · Ausgabe am 7.12.2008**  
**Abgabe am 10.12.2008 vor Beginn der Vorlesung**

## Aufgabe 1. (HMM für Jahreszeiten)

1. Erstelle ein HMM mit vier Zuständen entsprechend den vier Jahreszeiten, wobei jeder Zustand entweder **schneit** oder **nicht schneit** emittiert. Jeder Übergang zwischen zwei Zuständen soll einem Tag entsprechen.
2. Überlege dir eine vernünftige Belegung der Parameter für die Emission- und Transitionswahrscheinlichkeiten. Skizziere das Modell.
3. Benutze das Modell, um 100 Jahre a 365 Schritten zu simulieren. Das Ergebnis sind also 100 mit der Jahreszeit annotierte Sequenzen der Länge 365 von **schneit** und **nicht schneit**.
4. Benutze diese Sequenzen, um die Transitionswahrscheinlichkeiten neu zu bestimmen. Die Emissionswahrscheinlichkeiten bleiben wie vorher angenommen und müssen nicht erneut bestimmt werden.
5. Benutze Bootstrapping auf der Menge der Jahressequenzen, um eine Idee für die Varianz des Schätzers aus 4. zu bekommen. Ziehe mehrmals 100 Jahressequenzen mit Zurücklegen aus dem Ergebnis von 3. für das Bootstrapping.
6. Re-annotiere die Sequenzen mithilfe der neugeschätzten Transitionsmatrix (Mittelwerte über Bootstrap-Schätzer) und quantifiziere den Fehler.
7. Interpretiere sämtliche Ergebnisse.
8. Warum ist dieses HMM nicht so gut zur Darstellung, wann es im Verlauf eines Jahres schneit? Wie könnte man das HMM verbessern?