

23. November 2005

**Abgabe: 30. November 2005**

Die Aufgaben 1 und 2 beziehen sich auf das in der Vorlesung beschriebene Modell zum Finden von CpG-islands mit den versteckten Zuständen  $C^+$ ,  $G^+$ ,  $N^+$ ,  $C^-$ ,  $G^-$ ,  $N^-$ . Lassen Sie (insbesondere in Aufgabe 2) für die Emissionswahrscheinlichkeiten von A und T aus  $N^+$  und  $N^-$  auch andere Werte als  $\frac{1}{2}$  zu.

**Aufgabe 1:** Schreiben Sie ein Programm zum Simulieren von DNA-Sequenzen gemäß dem CpG-islands-HMM.

**Aufgabe 2:** Programmieren Sie den Baum-Welch-Algorithmus für das CpG-islands-HMM und erproben Sie Ihr Programm an simulierten Sequenzen aus Aufgabe 1 sowie einigen menschlichen DNA-Sequenzen aus den einschlägigen Datenbanken.

**Aufgabe 3:** Implementieren Sie den Viterbi-Algorithmus so, dass Ihr Programm als Eingabe jedes HMM mit einem endlichem Zustandsraum  $\mathcal{Z}$ , einem endlichem Emissions-Alphabet  $\mathcal{A}$  und dazugehörigen Übergangs- und Emissionswahrscheinlichkeiten akzeptiert. Erproben Sie dieses Programm an den Sequenzen aus Aufgabe 1 und Aufgabe 2 mit den dort vorausgesetzten bzw. gefundenen Wahrscheinlichkeiten.