Übungen Algorithmen der Bioinformatik II
Wintersemester 2005/06
Prof. Dr. Stefan Posch, Dr. Birgit Möller

Institut für Informatik
Universität Halle

Blatt 11

Aufgabe 11.1 Der Datensatz "sigma70_fg.txt" auf der Homepage zur Vorlesung besteht aus 238 Sigma-70 Bindungsstellen der Länge 12.

Der Datensatz "sigma70_bg.txt" ist ein zugehöriger Hintergrunddatensatz von ebenfalls 238 Sequenzen der Länge 12.

Teilen Sie beide Datensätze jeweils in zwei Hälften, wir erhalten so zwei TFBS-Datensätze F_1 und F_2 und zwei Hintergrunddatensätze B_1 und B_2.

(a) Konstruieren Sie zwei Bayes-Klassifikatoren K_i, indem Sie jeweils auf B_i und $F_i, i \in \{1, 2\}$, ein inhomogenes bzw. homogenes MM(0) für die TFBS bzw. den Hintergrund trainieren.

Machen Sie sich Gedanken zur Schätzung der a priori Wahrscheinlichkeiten der beiden Klassen!

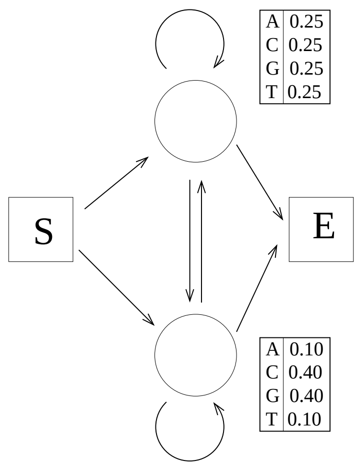
(b) Berechnen Sie für alle vier Markov-Modell jeweils das Sequenz-Logo.

(c) Wenden Sie beide Klassifikatoren auf beide Datensätze an und bestimmen Sie die Fehlerrate.

Aufgabe 11.2 Betrachten Sie das Beispiel-HMM aus der Vorlesung:

Die Übergangswahrscheinlichkeiten $C_{l,j}$ seinen:

		j			
l	S	0	0.5	0.5	0
	1	0	0.6	0.3	0.1
	2	0	0.3	0.6	0.1
	E	0	0	0	1



(a) Geben Sie formal ein äquivalentes HMM **ohne** Start- und Endzustand an.

(b) Wieviele mögliche Pfade können (für beide Varianten der HMM) zur Erzeugung der Sequenz $\vec{x}$ =`AGT` führen.

→ Bitte wenden!

Abgabe: 19.1.06

(c) Berechnen Sie jeweils $P(\pi, \vec{x}|\vec{\theta})$ sowie $P(\vec{x}|\vec{\theta})$.

(d) Wieviele mögliche Pfade führen zur Erzeugung einer (beliebigen) Sequenz $\vec{x}$ der Länge $L'(\vec{x})$.

Aufgabe 11.3 Wie müssen im Forward-Algorithmus die Startwerte $\alpha_j(0)$ für **nicht** emittierende Zustände $j \notin \mathcal{E}$ gewählt werden.

Verwenden Sie die in der Vorlesungen getroffenen Vorraussetzungen:

- jeder Zustand $j \notin \mathcal{E}$ hat je höchstens einen nicht-emittierenden Vorgänger und Nachfolger ($\notin \mathcal{E}$)
- es gibt keinen (gerichteten) Zyklus mit ausschließlich nicht-emittierenden Zuständen

---

Abgabe: 19.1.06