

30. November 2005

Abgabe: 7. Dezember 2005

Aufgabe 1: Besorgen Sie sich 5 möglichst verschiedene, lange biologische Sequenzen (Proteine, DNA, ...) und schätzen Sie für diese sowie für die 5 simulierten Sequenzen in der Datei `Markoffordnung.txt` die Markoffordnung. Erläutern Sie, wieso die dabei von Ihnen verwendete Methode sinnvoll ist.

Aufgabe 2: Geben Sie für alle $X, Y \in \{D, V, H\}$ die Wahrscheinlichkeit an, mit der beim Alignment Sampling (siehe Abschnitt 2.6.2 der Vorlesung) Y gewählt wird falls, zuletzt X gewählt wurde. Beweisen Sie die Korrektheit des Verfahrens.

Aufgabe 3: Beweisen Sie, dass man beim "Viterbi-Training mit Simulated Annealing" durch die Verwendung der mit $1/T_i$ potenzierten Parameter tatsächlich zufällige Pfade z_i gemäß der Wahrscheinlichkeit

$$\frac{Ws_{\theta_i}(z_i, \text{Daten})^{1/T_i}}{\sum_z Ws_{\theta_i}(z, \text{Daten})^{1/T_i}}$$

erzeugt.