Übungen Sequenz- und Expressionsdatenanalyse

Sommersemester 2004

Dr. Ivo Grosse

IPK, Gatersleben
Institut für Informatik,
Universität Halle

Blatt 6

Aufgabe 6.1 (4 Punkte) Modifizieren Sie den EM-Algorithmus für Motive aus Aufgabe 2.3 (bzw. 3.1) dahingehend, daß Sie die Parameter p_{iu_i} nicht mehr als im E-Schritt zu schätzende *interne* Parameter sondern als fest vorgegebene *externe* Parameter betrachten. Setzen Sie die externen Parameter $p_{iu_i} = 1/L_i$, und wenden Sie den modifizierten EM-Algorithmus auf die Sequenzen `pgrc-16.ipk-gatersleben.de/~grosse/Lehre/SEDA/SS04/Uebungen/Serie_2/seq_1.txt` und `pgrc-16.ipk-gatersleben.de/~grosse/Lehre/SEDA/SS04/Uebungen/Serie_2/seq_2.txt` aus Aufgabe 2.3 (bzw. 3.1) an. Definieren Sie Ihre Initialisierung und Ihr Abbruchkriterium, und plotten Sie die log-likelihood $\ln P(x|\theta)$ als Funktion der Iterationszeit t für 100 Läufe des EM Algorithmus.

Aufgabe 6.2 (2 Punkte) Beweisen Sie die Äquivalenz des rechts-links Markov Modells und des links-rechts Markov Modells.

Aufgabe 6.3 (4 Punkte) Der Datensatz `pgrc-16.ipk-gatersleben.de/~grosse/Lehre/SEDA/SS04/Uebungen/Serie_6/seq_1.txt` enthält ein Alignment von Donorstellen der Länge W bp. Berechnen Sie für $j = 2, 3, ..., W$ die Transinformation Y_1 und die χ^2-Teststatistik Y_2 als Maß der statistischen Abhängigkeit zwischen den Nukleotiden X_{j-1} und X_j, und plotten Sie Y_1 und Y_2 versus j. Unter der Annahme der Nullhypothese, daß es keine statistischen Abhängigkeeiten zwischen X_{j-1} und X_j gibt, ist sowohl Y_1 als auch Y_2 χ^2-verteilt mit 9 Freiheitsgraden. Für welche j finden Sie *statistisch signifikante* Abhängigkeiten, wenn Sie einen P-Wert kleiner als 0.05 als signifikant betrachten?

Wenden Sie diese Aqnalyse der statistischen Abhängigkeiten zwischen Nachbarnukleotiden auf den Datensatz `http://pgrc-16.ipk-gatersleben.de/~grosse/Lehre/SEDA/SS04/Uebungen/Serie_1/sigma_70.txt` aus Aufgabe 1.1 an. Für welche j finden Sie *statistisch signifikante* Abhängigkeiten, wenn Sie einen P-Wert kleiner als 0.05 als signifikant betrachten?

Abgabetermin: 19. Mai (in der Übung)