Übungen Algorithmen der Bioinformatik II
Wintersemester 2007/08
Prof. Dr. Stefan Posch, Dr. Birgit Möller

Institut für Informatik
Universität Halle

Blatt 10

Aufgabe 10.1

a) Plotten Sie die zweidimensionalen (verallgemeinerten) Gaussdichten

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2}+\frac{(y-\mu_y)^2}{\sigma_y^2}-2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right\}}$$

für die Werte von μ_x, μ_y, σ_x, σ_y und ρ aus der folgenden Tabelle:

μ_x	μ_y	σ_x	σ_y	ρ
0	0	1	1	0
3	0	1	1	0
0	-5	1	1	0
3	-5	1	1	0
0	0	4	1	0
0	0	1	$\frac{1}{9}$	0
0	0	4	$\frac{1}{9}$	0
0	0	4	1	0.2
0	0	4	1	0.4
0	0	4	1	0.6
0	0	4	1	0.8
0	0	4	1	1.0
0	0	4	$\frac{1}{9}$	0.4

Welchen Einfluss haben die einzelnen Parameter jeweils auf die Dichte?

b) Berechnen Sie die beiden Randdichten $f_X(x)$ und $f_Y(y)$ einer zweidimensionalen Gaussdichte

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2}+\frac{(y-\mu_y)^2}{\sigma_y^2}-2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right\}}$$

c) Berechnen Sie die bedingten Dichten $f_{X\,|\,Y}(x\,|\,y)$ und $f_{Y\,|\,X}(y\,|\,x)$ von $f(x,y)$.

d) Plotten Sie $f_X(x)$ und $f_{X\,|\,Y}(x\,|\,y)$ für $x \in [-10, 10]$ und $y \in \{-2, 0, 2, 4, 6\}$. Hier und in den folgenden beiden Beispielen sei $\mu_x = \mu_y = 0$, $\sigma_x = 1$, $\sigma_y = 2$ und $\rho = 0.75$.

Abgabe: 21.12.07

e) Plotten Sie $f_Y(y)$ und $f_{Y\,|\,X}(y\,|\,x)$ für $y \in [-10, 10]$ und $x \in \{-2, 0, 2, 4, 6\}$.

f) Plotten Sie $f(x, y)$ für $(x, y) \in [-10, 10] \times [-10, 10]$.

Aufgabe 10.2 Wir modellieren Sequenzen der Länge L über einem festen Alphabet der Größe D mit einem inhomogenen Markov Modell 1-ter Ordnung. Als Daten stehen uns N Sequenzen der Länge L zur Verfügung, die – wie üblich – als paarweise unabhängige Realisierung diese iMM(1) angenommen werden. Als a priori Verteilung der Parameter des iMM(1) nehmen wir statistisch unabhängige Dirichlet-Verteilungen an.

Geben Sie die explizit die ML-, MAP- und MP-Schätzwerte der Parameter des iMM(1) an.

Aufgabe 10.3 Wir betrachten N Sequenzen über einem festen Alphabet, die nun aber unterschiedliche Längen L_n haben.

Geben Sie für folgende Modelle für Sequenzen der Länge $L, L > 5$, jeweils die ML-Schätzer an:

(a) ein homogenes Markov Modell 5-ter Ordnung

(b) ein "periodisches" inhomogenes Markov Modell 5-ter Ordnung

(also ein iMM(5), für das gilt:
$P_l(X_l|X_{l-M}, \cdots, X_{l-1}) = P_k(X_k|X_{k-M}, \cdots, X_{k-1})$, falls $k = l\%5$)

Abgabe: 21.12.07