



Mustererkennung in der Bioinformatik

Dr. Hans A. Kestler, Dipl.-Inf. André Müller

WS 2003/2004

Übungsblatt 7

1. Aufgabe (16) Bayes-Klassifikator für Normalverteilungen

Gegeben seien Datenpaare $(\mathbf{x}_i, y_i) \in \mathbb{R}^d \times \{1 \dots m\}$, $i = 1 \dots n$, und wir nehmen an, dass die Datenpunkte unabhängig voneinander aus multivariaten Normalverteilungen gezogen wurden, mit den Wahrscheinlichkeitsdichten:

$$p(\mathbf{x}|\omega_j) = \frac{1}{(2\pi)^{d/2} \sqrt{|\mathbf{K}_j|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu_j)^T \mathbf{K}_j^{-1}(\mathbf{x} - \mu_j)\right)$$

Hierbei sind $\mu_j = E[\mathbf{x}|\omega_j]$ die klassenspezifischen Mittelwerte und $\mathbf{K}_j = E[(\mathbf{x} - \mu_j)(\mathbf{x} - \mu_j)^T|\omega_j]$ die Kovarianzmatrizen. $|\mathbf{K}_j|$ bezeichnet die Determinante der Matrix \mathbf{K}_j . Gegeben seien weiterhin die *a priori* Wahrscheinlichkeiten $P(\omega_j) = Pr\{Y = j\}$.

Der Bayes-Klassifikator h entscheidet sich bei einem vorliegenden \mathbf{x} für die wahrscheinlichste Klasse d.h. für

$$\begin{aligned} h(\mathbf{x}) &= \max_j Pr\{\omega_j|\mathbf{x}\} \\ &= \max_j p(\mathbf{x}|\omega_j)P(\omega_j) \end{aligned}$$

Implementieren Sie nun einen Klassifikator, der beim Training aus den übergebenen Datenpunkten die Mittelwerte und Kovarianzmatrizen berechnet und in der Lage ist für einen übergebenen \mathbf{x} -Vektor die wahrscheinlichste Klasse ω_j zu ermitteln. Die *a priori* Wahrscheinlichkeiten sollen als identisch vorausgesetzt werden d.h. $P(\omega_j) = 1/m$ (m ist die Anzahl der Klassen) für $j = 1 \dots m$.

Testen Sie den Klassifikator mit den beiden 2-dimensionalen Datensatzpaaren `b7-x0.txt` (Klasse 1), `b7-x1.txt` (Klasse 2) und `b7-2-x0.txt` (Klasse 1), `b7-2-x1.txt` (Klasse 2). Geben Sie jeweils Kontingenztafel und Fehlerrate für den trainierten Datensatz an und plotten Sie jeden Datensatz jeweils gemeinsam mit der entsprechenden Entscheidungsregion. Wie unterscheiden sich die Entscheidungsregionen?