Übungen Musterklassifikation
Sommersemester 2006

Abgabe am 30.6.06

Prof. Dr. Stefan Posch

Dipl.Inform. Andre Gohr

(andre.gohr@informatik.uni-halle.de)

Institut für Informatik
Universität Halle

Blatt 12

Aufgabe 12.1 (3 Punkte)
Vergleichen Sie

$$\vec{f}_o(\vec{x}) = E\{\vec{y}|\vec{x}\} = \int\limits_{\vec{y}} \vec{y} P(\vec{y}|\vec{x}) d\vec{y}$$

mit

$$\vec{d}(\vec{c}) = \begin{pmatrix} P(\omega_1|\vec{c}) \\ \vdots \\ P(\omega_k|\vec{c}) \end{pmatrix}$$

Aufgabe 12.2 (10 Punkte)
Ein 2-Klassen-Problem sei gegeben. Alle Mustervektoren einer Klasse wurden durch eine 2-D-Normalverteilung erzeugt. Der klassifizierte Gesamtdatensatz ist `2_class_mixture.dat` und ist auf der Website zur Übung zu finden.

(a) Zerlegen Sie den Gesamtdatensatz in einen *train*- und in einen *test*-Teildatensatz. Überlegen Sie sich, welche verhältnismäßige Aufteilung günstig ist und begründen Sie! (**1**)

(b) Schätzen Sie aus dem *train*-Datensatz die Parameter der 2-D-Normalverteilungen für jede Klasse. Geben Sie die geschätzten Parameter an! (**2**)

(c) Bauen Sie einen Likelihood-Klassifikator unter Verwendung der geschätzten Dichten der beiden 2-D-Normalverteilungen. Der Likelihood-Klassifikator entscheidet auf Klasse 1, wenn $\frac{P(\vec{c}\,|\,w_1)}{P(\vec{c}\,|\,w_2)} > \theta$ mit $\theta = 1$ gilt. Klassifizieren Sie den *test*-Teildatensatz und bestimmen Sie die erzielte Sensitivität für Klasse 1 (Sn_1) und 2 (Sn_2)! (**2**)

(d) Wie läßt sich das θ in obiger Verhältnisgleichung im Fall eines Bayes-Klassifikators interpretieren? (**1**)

(e) Nun soll in "kleinen Schritten" vom Likelihood- zum Bayes-Klassifikator übergegangen werden, indem θ beginnend bei 1 entsprechend Ihren Überlegungen aus der vorangegangenen Teilaufgabe so verändert wird, dass am Schluß bei entsprechendem θ der Klassifikator, der nach $\frac{P(\vec{c}\,|\,w_1)}{P(\vec{c}\,|\,w_2)} > \theta$ auf Klasse 1 entscheidet, sich wie ein Bayes-Klassifikator verhält (entscheidet). Bestimmen Sie fuer jeden Schritt wiederum die beiden klassenspezifischen Sensitivitäten! (**2**)

(f) Plotten Sie die erhaltenen Sensitivitäten: Sn_1 ueber $(1 - Sn_2)$! Interpretieren Sie den Plot! (**2**)