



Blatt 7

Aufgabe 7.1 (8 Punkte)

Die Stichprobe `2_class_mixture_train.dat` enthält eine Menge klassifizierter, 2-dimensionaler Merkmalsvektoren. Diese wurden vom einem 2-Klassen Zufallsprozess erzeugt, wobei die Merkmalsvektoren jeder Klasse durch eine 3-Komponenten-Misch-(2d)-Normalverteilung erzeugt wurden.

Ziel ist es, einen Normalverteilungs-Bayesklassifikator auf dieser Stichprobe zu trainieren. Schätzen Sie dazu die Parameter dieses Zufallprozesses! Verwenden Sie dabei die in der Vorlesung vorgestellten Vorgehensweisen:

- (a) harte Vektorquantisierung
- (b) weiche Vektorquantisierung

(als Abbruchkriterium könnte man die Anzahl der Iterationen verwenden) für je eine der zwei verschiedenen Annahmen:

- (I) klassenabhängige Dichten
Gehen Sie dabei von je einer 3-Komponenten-Misch-(2d)-Normalverteilung pro Klasse aus.
- (II) klassenunabhängige Dichten
Dazu schätzen Sie aus der Gesamtstichprobe eine 5-Komponenten-Misch-(2d)-Normalverteilung.

Für den Normalverteilungs-Bayesklassifikator sind zusätzlich die a_k^l zu bestimmen. Geben Sie für alle 4 verschiedenen Vorgehensweisen die Parametrisierungen (Mittelwerte, Kovarianzmatrix, a_k^l) an! Vergewissern Sie sich, dass die zu maximierende Zielfunktion H in allen 4 Fällen monoton wächst!

Aufgabe 7.2 (4 Punkte) Die Stichprobe `2_class_mixture_test.dat` enthält eine Menge nicht klassifizierter, 2-dimensionaler Merkmalsvektoren, die vom Zufallsprozess aus der vorangegangenen Aufgabe erzeugt wurden. Nutzen Sie die 4 verschiedenen, in der vorangegangenen Aufgabe bestimmten Normalverteilungs-Bayesklassifikatoren und klassifizieren Sie diese Merkmalsvektoren in die zwei Klassen (1 und 2). Welche

Unterschiede in den Klassifikationsergebnissen bezogen auf die 4 verschiedenen trainierten Bayesklassifikatoren stellen Sie fest? Geben Sie die klassifizierten Stichproben in eine Datei `2_class_mixture_test_classified_<methode>.dat` (`methode` codiert die 4 verschiedenen Vorgehensweisen) aus und schicken Sie diese als programmiertechnische Lösung mit!

Die Stichprobendateien finden Sie auf der Website zur Vorlesung.

Die programmiertechnischen Lösungen bitte rechtzeitig per mail schicken. Die theoretischen Lösungen bitte ausgedruckt oder handschriftlich am 26.5.06 im Raum 412 abgeben oder ebenfalls per mail schicken.