



Statistik III: Multivariate Stat. Verfahren

Übungsblatt 6

Bearbeitung: Do. 24.6.2004, 16.00 Uhr.

Alle Aufgaben können im Mac-Cip Pool bearbeitet werden.

1. Betrachtet werden wie in Aufgabe 4 von Blatt 5 die beiden Variablen *stearic* und *oleic* des **Olive Oils** Datensatz.
Parametrisieren Sie ein hinreichend feines Raster auf dem Wertebereich der beiden Variablen, und erstellen Sie für dieses Raster die Vorhersage bzgl. der LDA, so dass Sie eine Graphik der Entscheidungsregionen erhalten wie z.B. auf Seite 161 der Vorlesung.
2. Berechnen Sie nun für die Daten aus Aufgabe 1 die Quadratische Diskriminanz Analyse.
 - (a) Erstellen Sie die Konfusionsmatrix und vergleichen Sie die Ergebnisse mit denen der LDA.
 - (b) Visualisieren Sie die Entscheidungsregionen für die QDA wie in Aufgabe 1.
 - (c) Berechnen Sie jeweils die CCR (Correct Classification Rate) für die LDA und die QDA.
3.
 - (a) Parametrisieren einen 3-dimensionalen Würfel. Erstellen Sie für diese Daten ein Layout in 2 und 3 Dimensionen jeweils mit
 - i. `cmdscale`
 - ii. `sammon`, und
 - iii. `isoMDS`.Stellen Sie Ihre Ergebnisse graphisch dar, und verbinden Sie dabei benachbarte Punkte. Begründen Sie das Layout in 2-d, und beschreiben Sie die Unterschiede der drei Verfahren.
 - (b) Erstellen Sie ein MDS in 2 Dimensionen für die Variablen der 8 Fettsäuren des **Olive Oils** Datensatz mittels des Sammon Mappings. Interpretieren Sie die Ergebnisse! Wie wichtig ist die Normierung der Daten, und in wie weit spiegelt das Ergebnis die Korrelationen der Variablen wieder?
4. Berechnen Sie mittels `isoMDS` eine MDS in 2 und 3 Dimensionen für die normierten **Crabs** Daten.
 - (a) Benutzen Sie die Startkonfiguration der klassischen MDS.
 - (b) Geben Sie eine eigene Startkonfiguration vor:
 - i. Völlig zufällig
 - ii. Die zu erwartende Proximität der vier Gruppen vorgehend
 - (c) Vergleichen Sie die Ergebnisse besonders im Blick auf den Stress der verschiedenen Konfigurationen.
5. Betrachtet werden die **Crabs** Daten.
 - (a) Berechnen Sie aus den 5 Messgrößen 4 Cluster mittels `kmeans`, und überprüfen Sie die Ergebnisse graphisch.
 - (b) Nehmen Sie nun als Grundlage nicht die Daten selber sondern die 2. und 3. Hauptkomponente der 5 Messgrößen, und berechnen Sie erneut die 4 Cluster.
 - (c) Begründen Sie die beiden Ergebnisse aus (a) und (b)!