Lehrstuhl für Rechnerorientierte Statistik und Datenanalyse, Dr. Martin Theus
Institut für Mathematik
Universität Augsburg

Statistik III: Multivariate Stat. Verfahren

Übungsblatt 5

Bearbeitung: Do. 17.6.2004, 16.00 Uhr.
Alle Aufgaben können im Mac-Cip Pool bearbeitet werden.

1. Berechnen Sie für die folgenden Arbeitsplatz Daten eine kanonische Korrelation zwischen den Arbeitsplatz Charakteristika (Variablen 1–5) und der Zufriedenheit (Variablen 6–12). Gemessen wurden die Daten an $n = 784$ Beschäftigten einer Supermarktkette. Gegeben ist hier die Matrix der Korrelationen.

Feedback	1.00											
Bedeutsamkeit	0.49	1.00										
Vielfalt	0.53	0.57	1.00									
Identifizierung	0.49	0.46	0.48	1.00								
Selbständigkeit	0.51	0.53	0.57	0.57	1.00							
Vorgesetzten	0.33	0.30	0.31	0.24	0.38	1.00						
Karriereaussicht	0.32	0.21	0.23	0.22	0.32	0.43	1.00					
Finanziell	0.20	0.16	0.14	0.12	0.17	0.27	0.33	1.00				
Belastung	0.19	0.08	0.07	0.19	0.23	0.24	0.26	0.25	1.00			
Identifizierung	0.30	0.27	0.24	0.21	0.32	0.34	0.54	0.46	0.28	1.00		
Tätigkeit	0.37	0.35	0.37	0.29	0.36	0.37	0.32	0.29	0.30	0.35	1.00	
Allgemein	0.21	0.20	0.18	0.16	0.27	0.40	0.58	0.45	0.27	0.59	0.31	1.00

2. Betrachtet werden die Daten der **Australischen Krabben** der blauen Species (`sp==1`).
   (a) Berechnen Sie für die Variablen *FL* und *RW* die Projektionsgerade der optimalen Separation mittels LDA.
   (b) Plotten Sie die projizierten Dichten der beiden Gruppen mittels Dichteschätzern, und tragen Sie die Entscheidungsgrenze ein.
   (c) Vorausgesetzt wir nehmen eine Normalverteilung der beiden Gruppen an; wie groß ist die Wahrscheinlichkeit einer Fehlklassifikation?

3. Plotten Sie die Daten aus Aufgabe 2 in einem Scatterplot und differenzieren Sie dabei die beiden Gruppen mit verschiedenen Farben und Symbolen. Tragen Sie nun die 95% Konfidenz-Ellipsen ein für (i) die Gruppen individuell (ii) die Gruppen mit jeweils gepoolter Varianz (iii) für die Gesamtstichprobe.

4. Betrachtet werden die beiden Variablen *stearic* und *oleic* des **Olive Oils** Datensatz.
   (a) Plotten Sie die drei Regionen in verschiedenen Farben, und ergänzen Sie Konfidenzellipsen.
   (b) Berechnen Sie mittels der Function `lda` eine Lineare Diskriminanzanalyse für diese Variablen, und erstellen Sie die Konfusionsmatrix für die Vorhersage.

5. Leiten Sie die Entscheidungsregel der Quadratischen Diskriminanzanalyse im 1-dimensionalen Fall her. Was sind die Unterschiede zur LDA?