

Übungsblatt 14

Übung 14.1 Das bereitgestellte R-Script *scatter.matrices.r* stellt Funktionen bereit zum Zeichnen eines Streudiagramms der Datenwerte, zur Berechnung der Streumatrizen S_W und S_B sowie die beiden Separationsmaße

$$J_1 = \frac{\text{Spur}(S_B)}{\text{Spur}(S_W)} \quad J_2 = \text{Spur}(S_W^{-1} \cdot S_B)$$

Berechnen Sie damit die unterschiedlichen Werte für J_1 bzw. J_2 und erklären Sie die unterschiedlichen Werte anhand des Streudiagramms:

- a) Messpunkte Klasse ω_1 : (1,0), (1,3), (3,1), (3,3)
Messpunkte Klasse ω_2 : (2,2), (2,5), (4,2), (5,5)
- b) Messpunkte Klasse ω_1 : (1,0), (1,3), (3,1), (3,3)
Messpunkte Klasse ω_2 : (2,5), (4,4), (5,3), (5,5)
- c) Messpunkte Klasse ω_1 : (1,1), (1,3), (3,1), (3,3)
Messpunkte Klasse ω_2 : (0,0), (0,4), (4,0), (4,4)

Hinweis: Wie Sie die obigen Daten in R inline im Code definieren können und wie Sie die Funktionen J_1 , J_2 und *scatter.plot* aufrufen, können Sie dem kurzen Testcode am Ende des Scripts *scatter.matrices.r* entnehmen.

Übung 14.2 Der in R eingebauten *Iris-Datensatz* (Anderson, 1935) enthält für drei Iris-Sorten Messungen von Länge und Breite der Sepalen und Petalen.

Welche der sechs Kombination zweier Features führt zum höchsten Wert für J_1 oder J_2 bezüglich der Unterscheidung der Klassen *versicolor* und *virginica*? Wie sieht für diese Kombination das Streudiagramm aus?

Anleitung: Das erste und vierte Features aus dem Iris-Datensatz herausschneiden und das Streumaße J_1 berechnen können Sie mittels

```
features <- iris[,c(1,4)]
J1(features, iris$Species)
```

Sie können so die Qualitätsmaße J_1 und J_2 ausrechnen und in die beiden folgenden Tabellen eintragen:

J_1	s_len	s_wid	p_len	p_wid
s_len	X			
s_wid	X	X		
p_len	X	X	X	

J_2	s_len	s_wid	p_len	p_wid
s_len	X			
s_wid	X	X		
p_len	X	X	X	