

Übungsblatt 15

Übung 15.1 Bestimmen Sie für den Anderson's *Iris-Datensatz* (1935), für die Klassen *versicolor* und *virginica* und zwei Features Ihrer Wahl mittels PCA die Hauptkomponente und zeichnen Sie deren Richtung in ein Streudiagramm ein.

Hinweise:

- Die Klasse (Species) *setosa* können Sie entfernen mittels

```
noisetosa <- iris[iris$Species != "setosa",]  
noisetosa$Species <- factor(noisetosa$Species)
```

Die zweite Zeile ist erforderlich damit das Klassenlabel (R-Datentyp *factor*) die Information verliert, dass es die Klasse "setosa" mal gab.

- Das Streudiagramm können Sie wieder mit der Funktion *scatter.plot* aus der bereitgestellten Datei *scatter.matrices.r* zeichnen.
- Um die Richtung in das Streudiagramm einzuzeichnen, benutzen Sie einfach die Funktion *lines()*, wobei Sie nur zwei Punkte der Richtung angeben. Die beiden Punkte können Sie erhalten, indem Sie den Eigenvektor einmal zum Plot-Mittelpunkt addieren und einmal subtrahieren, z.B.

```
xm <- colMeans(features)  
x1 <- xm - eig$vector[,1]  
x2 <- xm + eig$vector[,1]  
lines(c(x1[1], x2[1]), c(x1[2], x2[2]), col="blue")
```

Übung 15.2 Bestimmen Sie die Hauptkomponente aller vier Features für Anderson's *Iris-Datensatz* für die Klassen *versicolor* und *virginica*. Decken sich die beiden wichtigsten Features mit der besten Zweier-Kombination aus Übung 14.2?