



Explorative Datenanalyse mit R und iPlots

Übungsblatt 6

Bearbeitung: Mi. 22. Juni 2005, 10.15 Uhr.

Alle Aufgaben können im Cip Pool bearbeitet werden.

1. Auf dem Flughafen in Newark, NJ (EWR) finden Sie auf einem Zettel folgenden R-Code:

```
#
# Fix iplot.opt() for ihist
#
.iplot.getPar <- function(plot) {
  #
  .jcall(plot$obj, "D", "getHistParam")[2]
  #
}
#
# Loop for interactive change of ihist bin-width
# (density curve is automatically updated!)
#
ilines(c(0,0), c(0,0), col="marked", visible=FALSE)
while (!is.null(ievent.wait())) {
  if (
```

Leider ist der weitere Code durch einen großen Flecken Pepsi unleserlich, und wir wissen nicht, was der unbekannte Author aus der tschechischen Republik da schreiben wollte.

Erstellen Sie mit Hilfe der oben angegebenen Funktion `.iplot.getPar` ein `ihist` mit eingezeichnetem Dichteschätzer, der auch bei Veränderung der Binbreite im `ihist` immer noch eine korrekte Dichteschätzung anzeigt.

2. Suchen Sie mittels iPlots für die Hurricane Daten möglichst homogene Untermengen bezüglich des Klassifikationsproblems. Wie lassen sich diese Untermengen motivieren bzw. erklären?
3. Erstellen Sie für die Variable Landfall Klassifikationsmodelle via logistischer Regression und Klassifikationsbäumen. Betrachten Sie:
 - (a) Die Klassifikationsgüte der beiden jeweils besten Modelle
 - (b) Den Unterschied zwischen der Klassifikationsgüte auf Test- und Trainingsdaten
 - (c) Für verschiedene Entscheidungsgrenzen $\kappa \in [0, 1]$ das Verhältnis zwischen "false Positives" und "false Negatives". Plotten Sie diese in einen Scatterplot, der als x -Achse den Anteil der "false Positives" und als y -Achse den Anteil der "false Negatives" plottet. Verbinden Sie die Punkte nach aufsteigendem κ .