Prof. Dr. U. Stadtmüller
M. Weyhmüller

Sommer 2012
30.04.2012
Blatt 02

# Übungen zur Funktionalen Datenanalyse

(Zu bearbeiten bis Montag, den 07.05.2012)

1. Lade die Datei `data02.dat` von der Homepage. Sie enthält (zufällige) Messzeitpunkte $x_i$ in der ersten, und zu diesen Zeitpunkten gemessene Werte $y_i$ in der zweiten Spalte ($1 \leq i \leq 200$). Man kann davon ausgehen, dass die Messwerte zufällige Messfehler $\varepsilon_i$ enthalten, die unabhängig und identisch verteilt sind, mit $\mathrm{E}(\varepsilon_i) = 0$.

   Passe zunächst ein lineares Modell an die Daten an. Verwende dann den Nadaraya-Watson-Schätzer mit einem Rechteckskern und versuche eine passende Bandbreite zu finden. Plotte die Punkte, sowie die geschätzten Regressionsfunktionen.

   *Hinweis*: Evtl. ist die Funktion `ksmooth()` hilfreich.

   (4 Punkte)

2. Der Datensatz `BostonHousing` enthält Ergebnisse einer Volkszählung aus dem Jahr 1970, zu 506 Gebieten der Stadt Boston. Der Datensatz ist im Paket `mlbench` enthalten und kann mit dem Befehl `data(BostonHousing)` verfügbar gemacht werden. Wir vermuten, dass der prozentuale Anteil an Bewohnern eines Bezirks, die der Unterschicht zugerechnet werden (Spalte `lstat`) den Immobilienpreis des Bezirks beeinflusst. Der Median des Hauspreises des Bezirks befindet sich im Datensatz in der Spalte `medv`.

   Passe zunächst ein lineares Modell an. Schätze dann die Regressionsfunktion mit dem Nadaraya-Watson-Schätzer, wie er in der Funktion `npreg` im Paket `np` implementiert ist. Verwende jeweils den Gleichverteilungs-, den Normalverteilungs- und den Epanechnikov-Kern. Verwende jeweils einmal die automatisch ermittelte Bandbreite und einmal eine Bandbreite Deiner Wahl. Plotte die Datenpunkte, sowie die Ergebnisse der Schätzungen.

   *Hinweis*: Der Kern kann in der Funktion `npreg()` mit der Option `ckertype="Name"` gewählt werden.

   (4 Punkte)

3. $X_1, \ldots, X_n$ sei eine Stichprobe zur Dichte $f \in C_2$. Die Dichte $f$ kann mit einem Kerndichteschätzer
$$\hat{f}_n(x;h) := \frac{1}{nh} \sum_{j=1}^{n} K\Big(\frac{x - X_j}{h}\Big)$$
   geschätzt werden. Dabei sei $K$ ein Kern der Ordnung $(0,2)$ und $h > 0$.

   (a) Zeige: $\mathrm{E}\Big(\frac{1}{h}K\big(\frac{x-X_1}{h}\big)\Big) = f(x) + f''(x) \int_{\mathbb{R}} \frac{h^2}{2} y^2 K(y)\,\mathrm{d}y + \mathrm{o}(h^2)$.

   (b) Zeige: $\mathrm{Var}(\hat{f}_n(x;h)) = \frac{1}{nh} f(x) \|K\|_2^2 + \mathcal{O}(\frac{1}{n})$

   (2 + 2 Punkte)