Prof. Dr. U. Stadtmüller
M. Weyhmüller

Sommer 2012
14.05.2012
Blatt 04

Übungen zur Funktionalen Datenanalyse

(Zu bearbeiten bis Montag, den 21.05.2012)

1. Betrachte erneut 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 ε_i enthalten, die unabhängig und identisch verteilt sind, mit $\mathbb{E}(\varepsilon_i) = 0$.

   Projekt A-VP: Schätze die Regressionsfunktion mit `loess()`. Plotte die Daten mit der Regressionsfunktion in ein gemeinsames Schaubild. Erstelle mittels Bootstrap (200 Durchläufe) einen Variability-Plot, wie in der Vorlesung beschrieben (Punktweise 2,5% und 97,5%-Quantil des Bootstrap-Verfahrens) für verschiedene Bandbreiten.

   Projekt B-CV: Implementiere das Leave-One-Out-Cross-Validation-Verfahren aus der Vorlesung zur Bestimmung der Bandbreite h. Verwende jeweils die Funktion `locpoly()` aus dem Paket `KernSmooth` um die Regression durchzuführen. Erstelle einen Plot verschiedener Werte von h und $CV(h)$ um das gefundene Optimum, sowie einen Plot der Daten mit der Regressionsfunktion. Verwende dabei die durch das Verfahren gefundene Bandbreite. Erstelle diese Plots für verschiedene Grade des Polynoms bei der Regression.

   Die Projekte sind bis Donnerstag, den 24.05.2012 zu bearbeiten.

   (8 Punkte)

2. Es sei K ein stückweise stetiger Kern der Ordnung $(0, 2)$ mit Träger $[-1, 1]$, $h < 1$ und

   $$s_{\mu,n}(x; h) = \frac{1}{nh} \sum_{\nu=1}^{n} (x - X_\nu)^\mu K\Big(\frac{x - X_\nu}{h}\Big)$$

   wie in der Vorlesung. Außerdem seien die Messpunkte X_i gemäß der Dichte $f_X \in C_3(\mathbb{R})$ verteilt (i.i.d.). Zeige, dass

   (a) $\mathbb{E}(s_{\mu,n}(x; h)) = h^2 c_2\big(f(x)\delta_{\mu,2} - f'(x)\delta_{\mu,1} + \frac{1}{2} f''(x)\delta_{\mu,0}\big) + \mathcal{O}(h^3)$

   (b) $\mathrm{Var}(s_{\mu,n}(x; h)) = h^{2\mu-1} \frac{1}{n} \int_{\mathbb{R}} u^{2\mu} K^2(u) f_X(x) \,\mathrm{d}u + \mathcal{O}\big(\frac{1}{n}(h^4 + h^{2\mu})\big)$

   Dabei sei $\delta_{a,b} = 1$ falls $a = b$ und 0 sonst. Wie in der Vorlesung ist $c_2 = \int_{\mathbb{R}} x^2 K(x) \,\mathrm{d}x$.

   (2 + 2 Punkte)