Arbeitsblatt 9

A 1 Bearbeiten Sie die Teilaufgaben e), f) und g) aus A1 von Blatt 8. Dazu sollten Sie den Begriff „effective degrees of freedom“ verstehen:

- Im linearen Modell gilt
$$\hat{y}_i = (Hy)_i$$
mit $H = X(X^TX)^{-1}X^T$. H ist eine symmetrische idempotente pxp-Matrix, deren Einträge nur von x abhängen. Die Anzahl an zu schätzenden Parametern im linearen Modell ist p = Spur(H).
- Für eine Smoothing Regression Spline aus den Daten $y = (y_1, \ldots y_n)$ und $x = (x_1, \ldots x_n)$ kann gezeigt werden, dass
$$\hat{y}_i = \hat{f}(x_i) = (S_\lambda y)_i.$$
Dabei ist S_λ eine nxn-Matrix, deren Einträge nur von λ und x abhängen. Im Smoothing Spline Modell definiert man analog die effektive Anzahl zu schätzender Parameter („effective degrees of freedom“) als Spur(S_λ).

Aus Spur(S_λ) lässt sich λ berechnen. So lässt sich z.B. zu einem polynomialen Modell mit p zu schätzenden Parametern eine Smoothing Spline gleicher Modellkomplexität definieren, indem λ so gewählt wird, dass Spur(S_λ)=p. In der Funktion smooth.spline wird über den Parameter df die effektive Anzahl zu schätzender Parameter übergeben, woraus dann λ berechnet wird.

A 2

a) Zeigen Sie, dass der Nadaraya-Watson Kernel Average Smoother mit fester Bandbreite λ und dem Tricube-Kern differenzierbar ist. (Hinweis : Verwenden Sie die Regel von L‘Hôspital).

b) Was kann gesagt werden, wenn statt dem Tricube-Kern der Epanechnikov-Kern verwendet wird?

c) Was kann gesagt werden, wenn der Epanechnikov-Kern mit variierender K-Nächste-Nachbarn-Bandbreite $h_\lambda(x)$ verwendet wird?

A 3 Die Funktionen locfit, locfit.raw und loess schätzen Kernel Average Smoother und lokale polynomiale Regressionsmodelle; locfit.raw erlaubt dabei auch andere Kerne als den tricube-Kern, der den Funktionen locfit und loess zugrundeliegt. Die Smoothingparameter nn (locfit), alpha (locfit.raw) und span (loess) definieren dabei die (von x abhängige) Bandbreite $h_\lambda(x)$. Dabei wird $h_\lambda(x)$ genau so definiert, dass für die $nn \cdot 100\%$ (bzw $alpha \cdot 100\%$ bzw. $span \cdot 100\%$) nächsten Beobachtungen um x

$$K_{h_\lambda(x)}(x, x_i) > 0$$

gilt, bzw. bei Wahl des Gauss-Kerns in locfit.raw gilt alpha $= 2 \cdot \sigma$.

Betrachten Sie den Datensatz ethanol (Package locfit), aus dem ein statistisches Modell $\hat{f}$ für $Y = NOx$ in Abhängigkeit von $X = E$ geschätzt werden soll.

a) Schätzen Sie $\hat{f}$ als K-Nächste-Nachbarn-Smoother jeweils mit Smoothingparameter 0.1, 0.3 und 0.5 sowie mit aus einer 5-fach-Kreuzvalidierung bestimmtem Smoothingparameter.

b) Schätzen Sie $\hat{f}$ auch als Kernel Average Smoother mit tricube-Kernel und jeweils den drei Smoothingparametern 0.1, 0.3 und 0.5 sowie mit dem aus einer 5-fach-Kreuzvalidierung bestimmtem Smoothingparameter. Plotten Sie die vier Kernel Average Smoother $\hat{f}$ in das Streudiagramm der beobachteten Daten.

c) Schätzen Sie nun $\hat{f}$ aus einer lokalen linearen Regression und einer lokalen quadratischen Regression. Wählen Sie 0.2 als Smoothing Parameter und plotten Sie die beiden Schätzer in das Streudiagramm. Welchen Wert schätzen Sie jeweils für $f(0.65)$ und $f(9)$? Wie scheinen sich im Allgemeinen Ihre beiden Schätzer $\hat{f}$ in den lokalen Minima und Maxima von f zu unterscheiden?

d) Wählen Sie nun den Smoothing-Parameter jeweils aus einer 5-fach-Kreuzvalidierung und plotten Sie auch diese beiden Schätzer in das Streudiagramm.

e) Nehmen Sie an, der wahre Zusammenhang zwischen X und Y ist lokal linear. Welcher der Schätzer $\hat{f}$ aus einem K-Nächste-Nachbarn-Smoother, einer lokalen linearen Regression und einer lokalen quadratischen Regression hat den kleinsten / größten Bias und welcher die kleinste / größte Varianz (bei fester Bandbreite K)? Begründen Sie Ihre Entscheidung.

f) Welches der Modelle liefert die kleinste/größte Quadratsumme der Residuen? Begründen Sie, inwieweit dies zu Ihren Erwartungen aus d) passt.

A 4 Es wird eine lokale lineare Regression mit fester Bandbreite λ und Gauss-Kern geschätzt, so dass in einem Punkt (x_j, y_j):

$$\hat{y}_j = \hat{f}(x_j) = \hat{b}_0 + \hat{b}_1 x_j \quad mit \quad (\hat{b}_0, \hat{b}_1) = \underset{b_0, b_1}{argmin} \sum_{i=1}^{n} K_\lambda(x_j, x_i)(y_i - b_0 - b_1 x_i)^2$$

Welchen Wert erreicht $\hat{y}_j$ asymptotisch mit $\lambda \to \infty$? Welchen Wert erreicht $\hat{y}_j$ asymptotisch mit $\lambda \to 0$?