Lösungen zu Übungsblatt 6

Aufgabe 1: Glättungsverfahren

(a) Wiederholung: Eigenschaften von B-Splines:

- k+1 Polynomabschnitte vom Grad k (Ordnung k+1)
- positive Funktionswerte der Basisfunktionen in einem durch k+2 Knoten aufgespannten Bereich, Null an allen anderen Stellen

$\Longrightarrow$ Form von B-Splines 1. Grades

$\Longrightarrow$ Form von B-Splines 2. Grades

(b) Differenzenpenalty 3. Ordnung (Herleitung):
P-Spline-Ansatz:

$$\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{m}\delta_j B_j(x_i)\right)^2 + \lambda \underbrace{\sum_{j=3+1}^{m}\left(\Delta^3\delta_j\right)^2}_{\text{Differenzenpenalty 3. Ordnung}} \xrightarrow{\underline{\delta}} \min$$

Andere Darstellung des Differenzen-Strafterms:

$$\sum_{j=4}^{m}\left(\Delta^3\delta_j\right)^2 = \left\|\underline{\underline{D}}_3\underline{\delta}\right\|^2 = \underline{\delta}^T\underline{\underline{D}}_3{}^T\underline{\underline{D}}_3\underline{\delta}$$

Differenzen 1. bis 3. Ordnung:

$$
\begin{array}{rcll}
\Delta^1\delta_j &=& \delta_j - \delta_{j-1} & \text{Differenz 1. Ordnung}\\
\Delta^2\delta_j &=& \Delta(\delta_j - \delta_{j-1}) & \text{Differenz 2. Ordnung}\\
&=& \delta_j - \delta_{j-1} - (\delta_{j-1} - \delta_{j-2}) &\\
&=& \delta_j - 2\delta_{j-1} + \delta_{j-2} &\\
\Delta^3\delta_j &=& \Delta(\delta_j - 2\delta_{j-1} + \delta_{j-2}) & \text{Differenz 3. Ordnung}\\
&=& (\delta_j - 2\delta_{j-1} + \delta_{j-2}) - (\delta_{j-1} - 2\delta_{j-2} + \delta_{j-3}) &\\
&=& \delta_j - 2\delta_{j-1} + \delta_{j-2} - \delta_{j-1} + 2\delta_{j-2} - \delta_{j-3} &\\
&=& \delta_j - 3\delta_{j-1} + 3\delta_{j-2} - \delta_{j-3} &
\end{array}
$$

$$
\Longrightarrow \left\| \underline{\underline{D}}_3 \underline{\delta} \right\|^2 = \left\| \begin{pmatrix} -1 & 3 & -3 & 1 & 0 & \dots & \dots \\ 0 & -1 & 3 & -3 & 1 & 0 & \dots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ \vdots & \dots & 0 & -1 & 3 & -3 & 1 \end{pmatrix} \begin{pmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_m \end{pmatrix} \right\|^2
$$

(c) Zum Output und der Graphik:

(i) Modellgleichung in Link-Form:

$$log(\mu) = \beta_0 + f(x_{\text{time}})$$

Modellgleichung in Responseform:

$$\mu = exp\{\beta_0 + f(x_{\text{time}})\} = exp\{\beta_0\} \cdot exp\{f(x_{\text{time}})\}$$

(ii) s(Zeit) ist höchst signifikant!

Sukzessive Interpretation des Outputs:

- Es wurde, wie bereits in (i) dargestellt, ein Poissonmodell mit natürlichem Link (d.h. Log-Link) gefittet.
- Das Modell besteht aus einem Intercept-Term und einem Spline. Die Parameterschätzer beider Komponenten sind höchst signifikant. Dabei ist auffällig, dass der P-Wert im Falle des Splines auf Basis eines χ^2-Tests (als quadratischer Version des Gauß-Tests) berechnet wurde. Es wurde ein kubischer Spline (`bs="cr"`) verwendet mit 8.886 effektiven Freiheitsgraden (`"edf"`) gefittet. Je niedriger die Anzahl effektiver Freiheitsgrade ist, umso glatter sieht die Splinekurve aus.
- R^2 und durch das Modell erklärte Devianz: Maße der Güte der Modellanpassung, die ein bisschen niedrig zu sein scheinen.
- UBRE-Score: Gütemaß, das anstelle des AICs berechnet wurde
- Scale est. : besagt, dass $\phi = 1$, also keine Über-/Unterdispersion angenommen wird.

- n=1337: Anzahl Beobachtungen (Hier: Tage, zu denen Daten vorliegen)

(iii) kurze Interpretation der Graphik:

- Fehlende Daten an manchen Tagen
- Deutlich erkennbare saisonale Schwankungen
- Insgesamt fallender Trend
- breitere Konfidenzbänder an den Rändern

(iv) Aus der Splinefunktion ablesen: $s(500) \sim 0.5$

$$\mu_{500} = exp(0.81508) \cdot exp(0.5) = 2.26 \cdot 1.65 = 3.7$$

Erwartete Zahl an Atemwegserkrankungen am Tag 500 liegt bei 3.7.