Gemischte Modelle Wintersemester 2015/2016

Vorlesung *Prof. Dr. Sonja Greven*
Übung *Sarah Brockhaus* **Thema 5**

Aufgabe 1

Zur Konstruktion von punktweisen Konfidenzbändern in einem additiven gemischten Modell.

(a) Denken Sie an die Eigenschaften des Schätzers von zufälligen Effekten. Warum ist $\mathbb{E}(\hat{\boldsymbol{b}}|\boldsymbol{b}) \neq \boldsymbol{b}$ und $\mathbb{E}(\hat{\boldsymbol{b}}) = \mathbb{E}(\boldsymbol{b}) = \boldsymbol{0}$?

(b) Sie wollen nun ein punktweises Konfidenzintervall für $f(z)$ berechnen, wobei $f(z) = \boldsymbol{C}_z \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{b} \end{pmatrix}$, mit $\boldsymbol{C}_z = (B_1(z), \ldots, B_d(z))$ und $B_j(\cdot)$ eine Basis-Funktion ist. Warum genügt es nicht, wie sonst üblich, die Varianz des Schätzers $\hat{f}(z)$ zur Konstruktion eines Konfidenzintervalls zu verwenden? Warum ist die quadratische Abweichung zwischen Schätzer und wahrem Wert (mean squared error, MSE), also $\{\hat{f}(z) - f(z)\}^2$ geeignet, um ein Konfidenzintervall zu bestimmen?

(c) Berechnen Sie folgenden Erwartungswert: $\mathbb{E}_{\boldsymbol{b}}\left[\mathbb{E}_{\boldsymbol{y}|\boldsymbol{b}}\left(\{\hat{f}(z) - f(z)\}^2|\boldsymbol{b}\right)\right]$.
Hinweis: Verwenden Sie den Satz vom iterierten Erwartungswert und nutzten Sie, dass Sie $\text{Cov}\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{b}} - \boldsymbol{b} \end{pmatrix} = (\boldsymbol{C}'\boldsymbol{R}^{-1}\boldsymbol{C} + \boldsymbol{A})^{-1}$ bereits kennen (vgl. Blatt 3 Aufgabe 1).

Aufgabe 2

Betrachten Sie den Datensatz *Soybean* aus dem R-Paket nlme.

(a) Lesen Sie die Hilfe zum Datensatz und wenden Sie die Funktionen `str()` und `plot()` auf *Soybean* an.

(b) Nutzten Sie die Funktion `xyplot()` aus dem R-Paket lattice, um eine Graphik des logarithmierten Gewichtes in Abhängigkeit von der Sorte (*Variety*) und der Parzelle (*Plot*) zu erstellen. Warum sollte man das logarithmierte Gewicht glatt und nicht linear modellieren?

(c) Fitten Sie ein Modell für das logarithmierte Gewicht mit Jahr, Sorte und glatter Funktion der Zeit (gleiche Form für alle Plots), sowie Plot-spezifischen zufälligen Intercepts. Lassen Sie sich auch eine geeignete Modellzusammenfassung ausgeben.
Hinweis: Verwenden Sie die Funktion `gamm()` aus dem R-Paket mgcv.

(d) Fügen Sie dem Modell aus (c) eine zufällige Steigung pro Plot hinzu.

(e) Schätzten Sie ein Modell für das logarithmierte Gewicht mit glatter Funktion der Zeit für jede Sorte, sowie einem festen Effekt für Jahr und zufälligen Intercepts für die Plots.