3 Generalisierte lineare Modelle

Aufgabe 3

Gegeben seien Beobachtungen $(y_i, \mathbf{x}_i), i = 1, \ldots, n$. Dabei sei $y_i \in \{0, 1\}$ eine binäre Responsevariable und $\mathbf{x}_i = (1, x_{i1}, \ldots, x_{ip})^T$ ein Vektor mit p Kovariablen. Der lineare Prädiktor ist durch $\eta = \mathbf{x}^T\boldsymbol{\beta}$ gegeben.

(a) Zunächst soll ein Probit-Modell gefittet werden. Leiten Sie die Score-Funktion $s(\boldsymbol{\beta})$ her.

(b) Die erwartete Fisher-Informationsmatrix $F(\boldsymbol{\beta})$ ist für GLMs allgemein gegeben durch

$$F(\boldsymbol{\beta}) = \sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i^T \left(\frac{\partial h(\eta_i)}{\partial \eta}\right)^2 / \sigma_i^2,$$

mit $\sigma_i^2 = \mathrm{var}(y_i)$.

Berechnen Sie mit Hilfe dieser Formel $F(\boldsymbol{\beta})$ für das Logit- und das Probit-Modell.

(c) Für die beobachtete Fisher-Informationsmatrix $F_{\mathrm{obs}}(\boldsymbol{\beta})$ erhält man im Logit-Modell

$$F_{\mathrm{obs}}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i^T \frac{\exp(\boldsymbol{x}_i^T\boldsymbol{\beta})}{(1+\exp(\boldsymbol{x}_i^T\boldsymbol{\beta}))}\left(1 - \frac{\exp(\boldsymbol{x}_i^T\boldsymbol{\beta})}{(1+\exp(\boldsymbol{x}_i^T\boldsymbol{\beta}))}\right).$$

Im Probit-Modell erhält man

$$\begin{aligned} F_{\mathrm{obs}}(\boldsymbol{\beta}) = & \sum_{i=1}^{n} \frac{\boldsymbol{x}_i\boldsymbol{x}_i^T}{\Phi^2(\boldsymbol{x}_i^T\boldsymbol{\beta})(1-\Phi(\boldsymbol{x}_i^T\boldsymbol{\beta}))^2}\Big(\phi'(\boldsymbol{x}_i^T\boldsymbol{\beta})\Phi^2(\boldsymbol{x}_i^T\boldsymbol{\beta})(1-\Phi(\boldsymbol{x}_i^T\boldsymbol{\beta})) + \phi^2(\boldsymbol{x}_i^T\boldsymbol{\beta})\Phi^2(\boldsymbol{x}_i^T\boldsymbol{\beta}) \\ & - y_i\left(\Phi(\boldsymbol{x}_i^T\boldsymbol{\beta})(1-\Phi(\boldsymbol{x}_i^T\boldsymbol{\beta}))\phi'(\boldsymbol{x}_i^T\boldsymbol{\beta}) - \phi^2(\boldsymbol{x}_i^T\boldsymbol{\beta})(1-2\Phi(\boldsymbol{x}_i^T\boldsymbol{\beta}))\right)\Big), \end{aligned}$$

wobei ϕ die Dichtefunktion und Φ die Verteilungsfunktion der N(0,1)-Verteilung bezeichnen.

Vergleichen Sie diese Formeln qualitativ mit Ihren Ergebnissen aus Teilaufgabe (b). Woraus ergeben sich diese Unterschiede?

(d) Zeigen Sie, dass in allen GLMs mit kanonischem Link $\frac{\partial h(\eta)}{\partial \eta} = \frac{\mathrm{var}(y)}{\phi}$ gilt. Welche weiteren Vorteile bietet die Verwendung des kanonischen Links im Vergleich zu anderen Linkfunktionen?

Aufgabe 4

Der Fisher-Scoring-Algorithmus zur Berechnung des ML-Schätzers des Parametervektors $\boldsymbol{\beta}$ ist in der Form der *iteratively weighted least squares* (IWLS), in der k-ten Iteration, gegeben durch

$$\hat{\boldsymbol{\beta}}^{(k+1)} = (\boldsymbol{X}^\top \boldsymbol{W}(\hat{\boldsymbol{\beta}}^{(\boldsymbol{k})})\boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{W}(\hat{\boldsymbol{\beta}}^{(\boldsymbol{k})})\tilde{\boldsymbol{\eta}}(\hat{\boldsymbol{\beta}}^{(\boldsymbol{k})}),$$

wobei $\boldsymbol{X}$ die Designmatrix inkl. Intercept, $\boldsymbol{W}(\boldsymbol{\beta})$ die Gewichtsmatrix und $\tilde{\boldsymbol{\eta}}(\boldsymbol{\beta}) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{D}(\boldsymbol{\beta})^{-1}(\boldsymbol{y} - \hat{\boldsymbol{\mu}})$ den Vektor der Pseudo-Beobachtungen bezeichnet.

(a) Implementieren Sie schrittweise den IWLS-Algorithmus für das Logit-Modell. Schreiben Sie dafür eine Funktion `iwls_logit()`, welcher der Response-Vektor $\boldsymbol{y}$, die Designmatrix $\boldsymbol{X}$ und ein Startwert für den Parametervektor $\boldsymbol{\beta}$ als Argumente übergeben werden. Als Abbruchkriterium bietet sich z.B. $\frac{||\hat{\boldsymbol{\beta}}^{(\boldsymbol{k+1})} - \hat{\boldsymbol{\beta}}^{(\boldsymbol{k})}||}{||\hat{\boldsymbol{\beta}}^{(\boldsymbol{k})}||} < 10^{-5}$ an.

Bestimmen Sie hierfür zunächst $\boldsymbol{D}(\boldsymbol{\beta})$ und $\boldsymbol{W}(\boldsymbol{\beta})$ für das Logit-Modell.

(b) Wenden Sie ihre Funktion aus Teilaufgabe (a) auf den Datensatz `shuttle` von Übungsblatt 2 an. Verwenden Sie die Startwerte $(0, 0)$, $(4, -0.1)$ und $(2, -2)$ für $\boldsymbol{\beta}$. Was fällt dabei auf? Vergleichen Sie Ihre Ergebnisse mit denen der Funktion `glm()` (Blatt 2/Aufgabe 1).