

Aufgabenblatt 5

Überblick

In einem *binären Wahlmodell* ist die abhängige Variable binär. Wir modellieren sie nicht direkt, sondern stattdessen die Wahrscheinlichkeit, dass sie den Wert 1 annimmt (dass sie den Wert 0 annimmt, folgt dann direkt aus der Gegenwahrscheinlichkeit). Lineare Modelle weisen hier Nachteile auf: Prognosen können negativ sein oder oberhalb von 1 liegen und es liegt zwangsläufig Heteroskedastizität vor. Als bessere Alternative gibt es *Logit-* und *Probit-Modelle*, sie werden mit der Maximum-Likelihood Methode geschätzt. Mit *ROC* Kurven können die Klassifizierungsgüten verschiedener Modelle verglichen werden.

Fortsetzung Aufgabenblatt 4

1. Angenommen, die abhängige Variable y ist nun binär. Welche der Annahmen im klassischen linearen Regressionsmodells gelten dann nicht mehr?
2. Welche neue Interpretation bekäme $\hat{\beta}_1$?

Aufgaben

1. Simulieren Sie Daten wie nachfolgend angegeben und schätzen Sie die Probit-Modelle

$$y = \beta_0 + \beta_1 x_1 + u \quad (1)$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \quad (2)$$

mit der Funktion `glm()`.

```
set.seed(1)
data <- MASS::mvrnorm(n = 50, mu = c(0, 0, 0), Sigma = diag(3))
data <- as.data.frame(data)
colnames(data) <- c("x_1", "x_2", "u")
beta_0 <- 1
beta_1 <- 0.2
beta_2 <- -0.5
y <- beta_0 + data$x_1 * beta_1 + data$x_2 * beta_2 + data$u
data$y <- ifelse(y >= 0.5, 1, 0)
```

2. Welches der Modelle (1) und (2) hat die “bessere” Null Deviance beziehungsweise Residual Deviance und was folgt daraus?
3. Bitte testen Sie für beide Modelle die Hypothese $H_0: \beta_k = 0, k = 1, \dots, K$ zu $\alpha = 10\%$.
4. Zeichnen Sie für beide Modelle die ROC Kurve mit der Funktion `pROC::roc()`. Welches Modell hat demnach die bessere Prognosequalität?