

Vaje 2: Splošni linearni model

Obravnavamo splošni linearni model

$$\mathbf{X} = \mathbf{Z}\beta + \epsilon,$$

kjer je $X \in \mathbb{R}^n$ vektor opazovanj, $\beta \in \mathbb{R}^d$ vektor parametrov, $Z \in \mathbb{R}^{n \times d}$ matrika konstant, $\epsilon \in \mathbb{R}^n$ vektor slučajnih odstopanj, za katerega vedno privzamemo $E(\epsilon) = 0$. Zapišemo $Z = SP$ za matriko $S \in \mathbb{R}^{n \times d}$ s paroma pravokotnimi stolpci, od katerih so nekateri lahko 0 (na stolpcih matrike Z izvajamo Gramm-Schmidtovo ortogonalizacijo). P je zgornjetrikotna matrika s pozitivni elementi na glavni diagonali. Tedaj je

$$Z^T Z = P^T J P,$$

kjer je $J = \text{diag}(j_1, \dots, j_d)$, $j_i = \|S^i\|^2$. Za reševanje enačbe $(Z^T Z)\hat{\beta} = Z^T \mathbf{x}$ konstruiramo konkretno rešitev

$$\hat{\beta} = (Z^T Z)' Z^T \mathbf{x},$$

kjer je $(Z^T Z)' = P^{-1} J P^{-T}$.

1. Za dano matriko Z in vektor X (datoteka primer1.csv) konstruiraj posplošeni inverz $(Z^T Z)'$ ter poišči oceno $\hat{\beta}$.

2. Linearna regresija

Predpostavimo linearni model

$$x_i = \beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + \varepsilon_i, \quad i = 1, \dots, n.$$

ε_i so neodvisne slučajne spremenljivke, vse porazdeljene po zakonu $N(0, \sigma^2)$. Vrednosti vektorjev X , Z_1 in Z_2 so v datoteki primer2.csv. Poišči oceno za $\hat{\beta}$ na tri načine:

- $\hat{\beta} = (Z^T Z)' Z^T \mathbf{x}$
- $\hat{\beta} = (Z^T Z)^{-1} Z^T \mathbf{x}$
- z uporabo R-ove funkcije `lm`

Ali se vrednosti ocen ujemajo?

3. ANOVA z enojno klasifikacijo

Obravnavamo m neodvisnih skupin, v vsaki od njih ($1 \leq j \leq m$) imamo n_j neodvisnih meritev, za katere privzamemo, da so porazdeljene kot $N(\mu + \alpha_j, \sigma_j^2)$. Mislimo si, da so α_j majhna števila. Pišimo $n = \sum_{j=1}^m n_j$ in naj bo

$$X_i = \mu + \alpha_j + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma_j^2) \quad \text{za} \quad \sum_{k=1}^{j-1} n_k + 1 \leq i \leq \sum_{k=1}^j n_k.$$

S primerno izbiro matrike Z , lahko zgornji model zapišemo kot splošni linearni model.

Naj bo $m = 4$, $\mu = 2$, $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (0.2, -0.5, 1, 2)$, $(\sigma_1, \sigma_2, \sigma_3, \sigma_4) = (0.05, 1, 0.03, 0.08)$ in $(n_1, n_2, n_3, n_4) = (25, 30, 25, 27)$. Simuliraj vrednosti X_i , konstruiraj primerno matriko Z ter preko posplošenega inverza matrike Z poišči ocene za $\mu, \alpha_1, \alpha_2, \alpha_3, \alpha_4$.