

Übung 5: Verallgemeinerte lineare Modelle 2 & Ridge-Regression und LASSO

Aufgabe 1

In einem verallgemeinerten linearen Modell ist die Nullhypothese

$$H_0 : \beta_j = 0 \quad \text{gegen} \quad H_1 : \beta_j \neq 0$$

mit Hilfe der Wald-Statistik zu testen. Für die ML-Schätzung von β_j und deren Varianz gilt

$$\hat{\beta}_j = 2 \quad \text{bzw.} \quad \text{Var}(\hat{\beta}_j) = 1,5.$$

Berechnen Sie den p -Wert und geben Sie an, ob die Nullhypothese bei einem Signifikanzniveau von $\alpha = 0,05$ zu verwerfen ist.

Aufgabe 2

Gegeben sei ein verallgemeinertes lineares Modell und zu testen sei die allgemeine lineare Hypothese

$$H_0 : \mathbf{C}\beta = \mathbf{d} \quad \text{gegen} \quad H_1 : \mathbf{C}\beta \neq \mathbf{d}$$

mit $\text{rang}(\mathbf{C}) = 2$. Für die Log-Likelihoodfunktion resultiert unter den beiden Hypothesen der Wert

$$\ln \left(L \left(\hat{\beta}_{H_0}; y_1, \dots, y_n \right) \right) = -1,3 \quad \text{bzw.} \quad \ln \left(L \left(\hat{\beta}_{H_1}; y_1, \dots, y_n \right) \right) = 2,3.$$

Prüfen Sie, ob die Nullhypothese H_0 zum Signifikanzniveau $\alpha = 0,05$ bzw. $\alpha = 0,01$ zu verwerfen ist.

Aufgabe 3

Bisher wurde in einem Unternehmen ein verallgemeinertes lineares Modell M_1 mit $k = 3$ erklärenden Variablen eingesetzt. Das Unternehmen steht vor der Entscheidung, ob zukünftig ein erweitertes Modell M_2 verwendet werden sollte, bei dem zu den bisherigen drei erklärenden Variablen zwei weitere hinzugefügt wurden. D.h zu testen ist

$$H_0 : M_1 \quad \text{gegen} \quad H_1 : M_2$$

Bei Vorliegen von $n = 20$ Beobachtungen betragen die skalierten Devianzen der beiden Modelle

$$\Delta_{M_1}^* = 7,2 \quad \text{bzw.} \quad \Delta_{M_2}^* = 1,1$$

- Entscheiden Sie anhand dieser Informationen, ob die Nullhypothese $H_0 : M_1$ zum Signifikanzniveau $\alpha = 0,05$ abzulehnen ist.
- Für die Log-Likelihoodfunktion des Modells M_2 hat man $\ln \left(L \left(\hat{\beta}_{M_2}; y_1, \dots, y_n \right) \right) = 5,4$ erhalten. Berechnen Sie den Wert der Log-Likelihoodfunktion von M_1 und des saturierten Modells.
- Wie würde die Entscheidung ausfallen, wenn M_1 kein Submodell von M_2 wäre?
- Wie groß müsste die Anzahl an Beobachtungen n mindestens sein, damit man sich im Aufgabenteil c) für das kleinere Modell entscheidet? Nehmen Sie dabei an, dass sich die Werte der Log-Likelihoodfunktion der beiden Modelle nicht verändern.

Aufgabe 4

Die Regressionskoeffizienten $\beta = (\beta_0, \beta_1, \dots, \beta_k)^T$ des linearen Regressionsmodells

$$y = X\beta + \varepsilon$$

sollen durch Lösen des Minimierungsproblems

$$\arg \min_{\beta \in \mathbb{R}^{k+1}} \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^k x_{ij} \beta_j \right)^2 \quad \text{unter der Nebenbedingung} \quad \sum_{j=1}^k |\beta_j| \leq t \quad (1)$$

geschätzt werden (LASSO-Schätzer). Der vorhandene Datensatz wird in einen Trainingsdatensatz vom Umfang n_1 und einen Validierungsdatensatz (Testdatensatz) vom Umfang $n_2 = n - n_1$ zerlegt. Erläutern Sie welche Auswirkung ein wachsender Wert von $t \geq 0$ auf die folgenden Größen hat:

- a) Geschätzter erwarteter quadrierter Prognosefehler für die Trainingsdaten, d.h.

$$\frac{1}{n_1} \sum_{i=1}^{n_1} (y_i - \hat{y}_i)^2$$

- b) Geschätzter erwarteter quadrierter Prognosefehler für die Testdaten, d.h.

$$\frac{1}{n_2} \sum_{i=n_1+1}^n (y_i - \hat{y}_i)^2$$

- c) Varianz, d.h. $\text{Var}(\hat{y})$

- d) Bias, d.h. $\mathbb{E}[\hat{y}] - \mathbb{E}[y]$

- e) Irreduzibler Prognosefehler, d.h. $\text{Var}(y)$

Aufgabe 5

Eine bekannte Eigenschaft der Ridge-Regression ist, dass Koeffizientenschätzungen für korrelierte erklärende Variablen tendenziell ähnlich sind, während bei LASSO teilweise sehr unterschiedliche Koeffizientenschätzungen resultieren können. Diese Eigenschaft soll im Folgenden anhand des sehr einfachen linearen Regressionsmodells

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

und $n = 2$ Beobachtungen untersucht werden. Dabei gelte weiter:

$$x_{11} = x_{12}$$

$$x_{21} = x_{22}$$

$$y_1 + y_2 = 0$$

$$x_{11} + x_{21} = 0$$

$$x_{12} + x_{22} = 0$$

- a) Zeigen Sie, dass mit der KQ-Methode, Ridge-Regression und LASSO für den Intercept β_0 jeweils die Schätzung $\hat{\beta}_0 = 0$ resultiert.
- b) Formulieren Sie das zur Ridge-Regression gehörende Optimierungsproblem.
- c) Zeigen Sie, dass im Falle der Ridge-Regression $\hat{\beta}_1 = \hat{\beta}_2$ gilt.
- d) Formulieren Sie das zu LASSO gehörende Optimierungsproblem.
- e) Zeigen Sie, dass im Falle von LASSO die Schätzungen $\hat{\beta}_1$ und $\hat{\beta}_2$ nicht eindeutig sind. Charakterisieren Sie die Lösungen.

Hinweis: Verwenden Sie hierzu die folgende alternative Formulierung des zu LASSO gehörenden Optimierungsproblems:

$$\min_{\beta \in \mathbb{R}^3} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2})^2 \quad \text{unter der Nebenbedingung} \quad |\beta_1| + |\beta_2| \leq t$$