Übung 2: Lineare Modelle für Regressionsprobleme

Aufgabe 1

Betrachtet wird ein polynomiales Regressionsmodell

$$t = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M + \varepsilon$$

mit $\mathbb{E}[\varepsilon] = 0$ und $\mathrm{Var}(\varepsilon) = \sigma^2 > 0$.

a) Stellen Sie das polynomiale Regressionsmodell als gewöhnliches lineares Modell mit Basisfunktionen dar.

b) Gegeben sei ein Trainingsdatensatz $\{(x_n, t_n) \in \mathbb{R}^2 : n = 1, \ldots, N\}$ mit $N \geq M + 1$ und für die Inputs gelte $x_i \neq x_j$ für $i \neq j$. Zeigen Sie, dass der KQ-Schätzer für $\mathbf{w} = (w_0, \ldots, w_M)^T$ eindeutig bestimmt ist.

   Hinweis: Die Determinante einer sog. $m \times m$-Vandermonde-Matrix

$$\mathbf{V} = \begin{pmatrix} 1 & \alpha_1 & \alpha_1^2 & \ldots & \alpha_1^{m-1} \\ 1 & \alpha_2 & \alpha_2^2 & \ldots & \alpha_2^{m-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_m & \alpha_m^2 & \ldots & \alpha_m^{m-1} \end{pmatrix}$$

   ist gegeben durch

$$\det(\mathbf{V}) = \prod_{1 \leq i < j \leq m} (\alpha_j - \alpha_i).$$

Aufgabe 2

Es werden die Annahmen a)-c) des klassischen linearen Modells vorausgesetzt (siehe Skript, Abschnitt 3.2) und der Ridge-Schätzer

$$\widehat{\mathbf{w}}^{Ridge} = \left(\mathbf{\Phi}^T \mathbf{\Phi} + \lambda \mathbf{E}\right)^{-1} \mathbf{\Phi}^T \mathbf{t}$$

betrachtet (siehe Skript, Abschnitte 3.5). Für die Designmatrix $\mathbf{\Phi} \in \mathbb{R}^{N \times (M+1)}$ wird ferner angenommen, dass $\mathbf{\Phi}^T \mathbf{\Phi} = N\mathbf{E}$ gilt. D.h. es wird angenommen, dass die Spalten von $\mathbf{\Phi}$ paarweise orthogonal sind.

a) Berechnen Sie den Erwartungswert $\widehat{\mathbf{w}}^{Ridge}$ und interpretieren Sie das Ergebnis.

b) Berechnen Sie die Varianz-Kovarianzmatrix von $\widehat{\mathbf{w}}^{Ridge}$ in Abhängigkeit von der Varianz-Kovarianzmatrix des KQ-Schätzers

$$\widehat{\mathbf{w}}^{KQ} = \left(\mathbf{\Phi}^T \mathbf{\Phi}\right)^{-1} \mathbf{\Phi}^T \mathbf{t}$$

   (siehe Skript, Abschnitte 3.2).

c) Weisen Sie mit Hilfe des Ergebnisses von Aufgabenteil b) nach, dass für die Varianzen der $M+1$ Komponenten von $\widehat{\mathbf{w}}^{Ridge}$ und $\widehat{\mathbf{w}}^{KQ}$

$$\mathrm{Var}(\widehat{w}_j^{Ridge}) < \mathrm{Var}(\widehat{w}_j^{KQ}) \quad \text{für } j = 0, \ldots, M$$

   gilt. Interpretieren Sie dieses Ergebnis zusammen mit dem Ergebnis aus Aufgabenteil a).

Aufgabe 3

Betrachtet wird das lineare Regressionsmodell

$$t = y(\mathbf{x}, \mathbf{w}) + \varepsilon \qquad \text{mit} \qquad y(\mathbf{x}, \mathbf{w}) = w_0 + w_1 x_1 + \ldots + w_M x_M,$$

den Inputs $\mathbf{x} = (x_1, \ldots, x_M)^T$, den Regressionskoeffizienten $\mathbf{w} = (w_0, w_1, \ldots, w_M)^T$ und dem zufälligem Fehler ε mit $\mathbb{E}[\varepsilon] = 0$ und $\mathrm{Var}(\varepsilon) = \sigma^2$. Die Regressionskoeffizienten sollen durch Lösen des Minimierungsproblems

$$\arg\min_{\mathbf{w} \in \mathbb{R}^{M+1}} \sum_{n=1}^{N} \left(t_n - w_0 - \sum_{j=1}^{M} x_{nj} w_j \right)^2 \quad \text{unter der Nebenbedingung} \quad \sum_{j=1}^{M} |w_j| \leq \eta$$

geschätzt werden (LASSO-Schätzer). Der vorhandene Datensatz wird in einen Trainingsdatensatz vom Umfang N_1 zur Schätzung von $\mathbf{w}$ und einen Testdaten- satz vom Umfang $N_2 = N - N_1$ zerlegt. Erläutern Sie welche Auswirkung ein wachsender Wert von $\eta \geq 0$ auf die folgenden Größen hat:

a) Empirisches Risiko für die Trainingsdaten, d.h.

$$\frac{1}{N_1}\sum_{n=1}^{N_1}\big(t_n - y(\mathbf{x}_n,\widehat{\mathbf{w}})\big)^2$$

b) Empirisches Risiko für die Testdaten, d.h.

$$\frac{1}{N_2}\sum_{n=N_1+1}^{N}\big(t_n - y(\mathbf{x}_n,\widehat{\mathbf{w}})\big)^2$$

c) Varianz, d.h. $\mathrm{Var}\big(y(\mathbf{x},\widehat{\mathbf{w}})\big)$

d) Bias, d.h. $y(\mathbf{x},\mathbf{w}) - \mathbb{E}\big[y(\mathbf{x},\widehat{\mathbf{w}})\big]$

e) Irreduzibler Prognosefehler, d.h. $\mathrm{Var}(\varepsilon) = \sigma^2$