Maschinelles Lernen aus Stochastischer Sicht — Sommersemester 2024

Vorlesung: Prof. Dr. Thorsten Schmidt
Übung: Moritz Ritter

Übungsblatt 1

Abgabe: Freitag, 26.04.2022, 18:00 Uhr.

Aufgabe 1 (4 Punkte)**.** Sei $(\Omega, \mathscr{F}) = (\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ für ein $n \in \mathbb{N}$. Sei $\Theta = \tilde{\Theta} \times (0, \infty)$ mit $\tilde{\Theta} = \mathbb{R}^d$ für ein $d \in \mathbb{N}$ mit $d \leq n$ und

$$P_\theta = \mathcal{N}(A\beta, \sigma^2 \operatorname{Id}),\ \ \theta = (\beta, \sigma^2) \in \Theta$$

mit einer Matrix $A \in \mathbb{R}^{n \times d}$, sodass $\operatorname{rg}(A) = d$. Das statistische Modell $(\Omega, \mathscr{F}, \{P_\theta : \theta \in \Theta\})$ heißt lineares Modell mit koordinatengebundener Darstellung. Für die einfache lineare Regression $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, $i = 1, \ldots, n$ haben A und die Parameter (β, σ^2) die Gestalt

$$A = (1\ X) \in \mathbb{R}^{n \times 2} \text{ und } \beta = (\beta_0, \beta_1) \in \mathbb{R}^2, \sigma^2 \in (0, \infty).$$

Finden Sie den Parameterraum Θ und die zugehörige Matrix A zu folgenden Modellen:

i) Polynomiale Regression:
$Y_i = \beta_0 + \beta_1 X_i + \cdots + \beta_k X_i^k + \varepsilon_i$.

ii) Mehrfache lineare Regression:
$Y_i = \beta_0 + \beta_1 X_{1,i} + \cdots + \beta_d X_{d,i} + \varepsilon_i$.

iii) Mehrfache polynomiale Regression:
$Y_i = \beta_0 + \beta_{1,1} X_{1,i} + \cdots + \beta_{1,k} X_{1,i}^k + \cdots + \beta_{d,1} X_{d,i} + \cdots + \beta_{d,k} X_{d,i}^k + \varepsilon_i$.

Aufgabe 2 (4 Punkte)**.** Sei $d \leq n$. Zeigen Sie, dass für eine Matrix $A \in \mathbb{R}^{n \times d}$ mit vollem Rang, $\operatorname{rg}(A) = d$, die Matrix $A^\top A \in \mathbb{R}^{d \times d}$ positiv definit und damit invertierbar ist.

Aufgabe 3 (4 Punkte)**.** Die Moore-Penrose-Inverse von $A \in \mathbb{R}^{n \times d}$ ist definiert durch

$$A^\dagger := (A^\top A)^{-1} A^\top \in \mathbb{R}^{d \times n}.$$

Weiter sei $\phi(x) := Ax$ und $W := \phi(\mathbb{R}^d) \subseteq \mathbb{R}^n$. Zeigen Sie folgende Aussagen:

1. Es gilt für eine Abbildung $\psi \colon \mathbb{R}^n \to \mathbb{R}^n$ die Äquivalenz

$$\langle x - \psi(x), y \rangle = 0 \text{ für alle } y \in W \quad \Longleftrightarrow \quad A^\top(x - \psi(x)) = 0.$$

2. Die Funktion $\psi(x) = AA^\dagger x$ erfüllt $A^\top(x - \psi(x)) = 0$.
3. Der kleinsten Quadrate Schätzer $\hat{\beta}$ für β, das heißt

$$\|X - A\hat{\beta}\| = \min_{\beta \in \mathbb{R}^d} \|X - A\beta\|,$$

ist gegeben durch $\hat{\beta} = A^\dagger X$.

Hinweis: Verwenden Sie für den dritten Teil den Projektionssatz: Sei $W \subset \mathbb{R}^n$ ein Unterraum von $\mathbb{R}^n$ und sei $x_0 \in \mathbb{R}^n$. Dann sind für $x \in W$ äquivalent: $\|x_0 - x\| = \inf_{y \in W} \|x_0 - y\| \Longleftrightarrow \langle x_0 - x, y \rangle = 0$ für alle $y \in W$.