EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Mathematisch-Naturwissenschaftliche Fakultät

Fachbereich Mathematik

Prof. Dr. Andreas Prohl
Armin Beck

Statistisches Lernen 1

Sommer-Semester 2024 **Tübingen, 15.04.2024**

Blatt 1

Problem 1. Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$, betrachte die Zerlegung

$$\Omega = \bigcup_{n\in\mathbb{N}} \Omega_n, \text{ wobei } \Omega_n \in \mathcal{F} \text{ und } \mathbb{P}[\Omega_n] > 0 \ \ (n \in N).$$

Sei $\mathcal{G} \subset \mathcal{F}$ die Unter-σ-Algebra, die von der Familie $\{\Omega_n\}_{n\in\mathbb{N}}$ erzeugt wird, d.h.,

$$G = \Big\{ \bigcup_{n\in\mathbb{N}} \Omega_n^{\epsilon_n} : \epsilon_1, \epsilon_2, ... \in \{0,1\} \Big\}, \text{ wobei } \Omega_n^1 = \Omega_n \text{ und } \Omega_n^0 = \emptyset.$$

Sei nun $X : \Omega \to \mathbb{R}^d$ eine $\mathcal{F}$-messbare Zufallsvariable auf Ω mit $\mathbb{E}[\|X\|] < \infty$. Zeige, dass

$$\mathbb{E}[X|\mathcal{G}](\omega) = \frac{\mathbb{E}[X \mathbb{1}_{\Omega_n}]}{\mathbb{P}[\Omega_n]}$$

für alle $\omega \in \Omega_n$ gilt.

Problem 2. Für einen gegebenen Datensatz, $D_n = \{(X_i, Y_i)\}_{i=1}^n$ ist lineare Regression ein Beispiel für einen Parameterschätzer m_n der Regressionsfunktion m. Der $\mathbb{R}^{1+d}$-wertige (zufällige) Parameter $\boldsymbol{a} \equiv (a_0, \underline{\boldsymbol{a}})^T$ bestimmt

$$m_n \equiv m_n(\cdot; \boldsymbol{a}, D_n) : x \mapsto a_0 + \langle \underline{\boldsymbol{a}}, x \rangle_{\mathbb{R}^d},$$

und der gesuchte optimale Parameter a^* minimiert das quadratische Funktional

$$\frac{1}{n} \sum_{i=1}^{n} |Y_i - (a_0 + \langle \underline{\boldsymbol{a}}, X_i \rangle_{\mathbb{R}^d}|^2.$$

Disskutiere wie der Parameter a^* berechnet wird.

Problem 3. Sei $D_n = \{(X_j, Y_j)\}_{j=1}^n$ ein Datensatz, und $m_n \equiv m_n(\cdot; D_n)$ ein Regressionsschätzer. Zeige

$$\mathbb{E}[|m_n(X) - Y|^2 | D_n] = \int_{\mathbb{R}^d} |m_n(x) - m(x)|^2 \mu[dx] + \mathbb{E}[|m(X) - Y|^2].$$

Date of Submission: 22.04.2024 12 Uhr (in Briefkasten).