EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Mathematisch-Naturwissenschaftliche Fakultät

Fachbereich Mathematik

Prof. Dr. Andreas Prohl
Armin Beck

Statistical Learning 1

Summer semester 2024 **Tübingen, 17.06.2024**

Assignment 9

Problem 1
Let m_n be the partitioning estimator. For a cubic partition with side length h_n assume that

a) $\mathcal{S} := \operatorname{supp}(\mathbb{P}_X)$ is bounded,

b) $\operatorname{Var}[Y|X=x] \leq \sigma^2$ for all $x \in \mathbb{R}^d$,

c) $|m(x) - m(z)| \leq C_{Lip}\|x-z\|$, for all $x, z \in \mathbb{R}^d$.

Then there is a constant $c_1 \equiv c_1(d, \operatorname{diam}(\mathcal{S}))$ such that

$$\mathbb{E}\Big[\int_{\mathbb{R}^d} |m_n(x) - m(x)|^2 \mu[dx]\Big] \leq c_1 \frac{\sigma^2 + \sup_{z\in\mathcal{S}} |m(z)|^2}{nh_n^d} + dC_{Lip}^2 h_n^2.$$

Hint:

a) In a first step, show

$$\hat{m}_n(x) := \mathbb{E}[m_n(x)|X_1, ..., X_n] = \frac{\sum_{i=1}^n m(X_i)\mathbb{1}_{\{X_i \in \mathcal{A}_n(X)\}}}{n\mu_n[\mathcal{A}_n(X)]} \tag{1}$$

and convince yourself that a corresponding variance-bias decomposition holds.

b) Then independently bound the appearing variance term and the bias term, where $n \cdot \mu_n[\mathscr{A}_{n,j}] \sim \mathscr{B}(n, \mu[\mathscr{A}_{n,j}])$ is used for the first term, and Jensen's inequality is applied for the latter.

Problem 2
In the lecture, we formulated and verified for the kernel estimator the following $\mathbb{L}^2$-estimate

$$\mathbb{E}\Big[\int_{\mathbb{R}^d} |m_n(x) - m(x)|^2 \mu[dx]\Big] \leq c_1 \frac{\sigma^2 + \sup_{z\in\mathcal{S}} |m(z)|^2}{nh_n^d} + C_{H\ddot{o}l}^2 h_n^{2p} =: I,$$

for constants $C_{H\ddot{o}l} \geq 0$, $p \in (0,1]$ and $\sigma^2 \geq 0$, provided that some boundedness respectively regularity assumptions hold for (X, Y) respectively m (similar as in Problem 1).
The upper bound I consists of two terms: verify that an optimal balancing of both suggests the choice

$$h_n = \Big(\frac{dc_1(\sigma^2 + \sup_{z\in\mathcal{S}} |m(z)|^2)}{2pC_{H\ddot{o}l}^2 n}\Big)^{\frac{1}{2p+d}},$$

such that for some $\bar{c} > 0$,

$$I \leq \bar{c}\Big(\frac{\sigma^2 + \sup_{z\in\mathcal{S}} |m(z)|^2}{n}\Big)^{\frac{2p}{2p+d}} C_{H\ddot{o}l}^{\frac{2d}{2p+d}}.$$

Date of Submission: 24.06.2024 in the mailbox at 12 noon.