EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Mathematisch-Naturwissenschaftliche Fakultät

Fachbereich Mathematik

Prof. Dr. Andreas Prohl
Armin Beck

Statistical Learning 1

Summer semester 2024 **Tübingen, 1.07.2024**

Assignment 11

Problem 1
Let $X_1, ..., X_n$ be independent, real-valued random variables. Let $a, b \in \mathbb{R}$ such that $a < b$ and assume $X_i \in [a, b]$ with probability one for all $i = 1, ..., n$. Let

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} \mathrm{Var}[X_i] > 0.$$

Show for all $\epsilon > 0$.

$$\mathbb{P}\Big[\Big|\frac{1}{n} \sum_{i=1}^{n} (X_i - \mathbb{E}[X_i])\Big| \leq 2 \exp\Big(- \frac{n\epsilon^2}{2\sigma^2 + \frac{2\epsilon}{3}(b-a)}\Big).$$

Problem 2
The splitting (regression) algorithm splits the sample set $D_n = D_{n_L} \dot{\cup} D_{n_T}$ to compute the regression estimator $m_n \equiv m_{n_L}^{(\hat{h})}(D_{n_L})$, with data-dependent parameter $\hat{h} \equiv \hat{h}(D_{n_T}; D_{n_L})$. Let $L > 0$ and T_L the truncation operator

$$T_L u := \begin{cases} u & \text{if } |u| \leq L \\ L \,\mathrm{sgn}(u) & \text{otherwise.} \end{cases}$$

Assume the distribution of (X, Y) satisfies $|Y| \leq L$ and $\max_{h \in P_n} \|m_{n_L}^{(h)}\|_\infty \leq L$ $\mathbb{P}$ a.s.. Let

$$\sigma_h^2 := \mathrm{Var}\left[|m_{n_L}^{(h)}(X) - Y|^2 - |m(X) - Y|^2 | D_{n_L}\right].$$

Show

$$\sigma_h^2 \leq 16 L^2 \Big(\mathbb{E}\big[|m_{n_L}^{(h)}(X) - Y|^2 | D_{n_L}\big] - \mathbb{E}[|m(X) - Y|^2] \Big)$$

Date of Submission: 8.07.2024 in the mailbox at 12 noon.