Dr. Stefan Richter
Statistische Analyse von maschinellen Lernalgorithmen
Wintersemester 2020/2021

3. Übungsblatt

Aufgabe 1 (Diskussion: Übertragung der Beweistechnik von Satz 3.12)
Lösung:

(a) Es gilt $\hat{R}_n(\hat{\beta}) \le \hat{R}_n(\beta^*)$, daher

$$R(\hat{\beta}) - R(\beta^*) = \underbrace{\hat{R}_n(\hat{\beta}) - \hat{R}_n(\beta^*)}_{\le 0} - \big\{\hat{R}_n(\hat{\beta}) - R(\hat{\beta}) - (\hat{R}_n(\beta^*) - R(\beta^*))\big\}.$$

(b) Es gilt $\tilde{\beta} - \beta^* = T(\hat{\beta} - \beta^*)$, daher

$$\|\Sigma^{1/2}(\tilde{\beta} - \beta^*)\|_2 = T \cdot \|\Sigma^{1/2}(\hat{\beta} - \beta^*)\|_2 \le \gamma.$$

Wegen Konvexität von $\beta \mapsto \hat{R}_n(\beta)$ gilt außerdem

$$\hat{R}_n(\tilde{\beta}) \le T \underbrace{\hat{R}_n(\hat{\beta})}_{\le \hat{R}_n(\beta^*)} + (1-T)\hat{R}_n(\beta^*) \le \hat{R}_n(\beta^*),$$

d.h. die Aussage aus (a) gilt auch für $\tilde{\beta}$.
Wir erhalten

$$\begin{aligned} R(\tilde{\beta}) - R(\beta^*) &\le \big|\hat{R}_n(\tilde{\beta}) - R(\tilde{\beta}) - (\hat{R}_n(\beta^*) - R(\beta^*))\big| \\ &\le \sup_{\beta : \|\Sigma^{1/2}(\beta - \beta^*)\|_2 \le \gamma} \big|(\hat{R}_n(\beta) - R(\beta)) - (\hat{R}_n(\beta^*) - R(\beta^*))\big|. \end{aligned}$$

(c) Auf A gilt

$$\|\Sigma^{1/2}(\tilde{\beta} - \beta^*)\|_2^2 = R(\tilde{\beta}) - R(\beta^*) \le Z_\gamma \le a\gamma,$$

daher

$$\|\Sigma^{1/2}(\tilde{\beta} - \beta^*)\|_2 \le (a\gamma)^{1/2} \le \frac{\gamma}{2}.$$

Nutzung $T \cdot \|\Sigma^{1/2}(\hat{\beta} - \beta^*)\|_2 \overset{s.o.}{=} \|\Sigma^{1/2}(\tilde{\beta} - \beta^*)\|_2$ liefert

$$\frac{\gamma \|\Sigma^{1/2}(\hat{\beta} - \beta^*)\|_2}{\gamma + \|\Sigma^{1/2}(\hat{\beta} - \beta^*)\|_2} \le \frac{\gamma}{2}$$

Umstellen $\Rightarrow \|\Sigma^{1/2}(\hat{\beta} - \beta^*)\|_2 \le \gamma$.
Wiederholung des Beweises für $\hat{\beta}$ statt $\tilde{\beta}$ liefert auf A:

$$R(\hat{\beta}) - R(\beta^*) \le a\gamma.$$

(d) Wir haben gesehen: $A \subset \{R(\hat{\beta}) - R(\beta^*) \le a\gamma\}$.
Gegenereignis:

$$\mathbb{P}(R(\hat{\beta}) - R(\beta^*) > a\gamma) \le \mathbb{P}(A^c).$$

(e) Elementare Rechnung:

$$\hat{R}_n(\beta) = \frac{1}{n}\| \underbrace{\mathbb{Y}}_{=\mathbb{e}+\mathbb{X}\beta^*} - \mathbb{X}\beta\|_2^2 = \frac{1}{n}\|\mathbb{e}\|_2^2 + \frac{2}{n}\mathbb{e}^T\mathbb{X}(\beta^* - \beta) + \underbrace{\frac{1}{n}\|\mathbb{X}(\beta^* - \beta)\|_2^2}_{\overset{X_i \text{ determ.}}{=} \|\Sigma^{1/2}(\beta - \beta^*)\|_2^2}.$$

$\Rightarrow$

$$\hat{R}_n(\beta) - R(\beta) = \frac{1}{n}\|\mathbb{e}\|_2^2 + \frac{2}{n}\mathbb{e}^T\mathbb{X}(\beta^* - \beta).$$

$\Rightarrow$

$$(\hat{R}_n(\beta) - R(\beta)) - (\hat{R}_n(\beta^*) - R(\beta^*)) = \frac{2}{n}\mathbb{e}^T\mathbb{X}(\beta^* - \beta).$$

Damit

$$\begin{aligned}
|Z_\gamma| &\leq \frac{2}{n}\sup|\mathbb{e}^T\mathbb{X}(\beta^* - \beta)| = \frac{2}{n}\sup|\mathbb{e}^T\mathbb{X}\Sigma^{-1/2}\Sigma^{1/2}(\beta^* - \beta)| \\
&\leq \frac{2}{n}\|\mathbb{e}^T\mathbb{X}\Sigma^{-1/2}\|_2 \sup\|\Sigma^{1/2}(\beta^* - \beta)\|_2 \\
&\leq \frac{2\gamma}{n}\|\mathbb{e}^T\mathbb{X}\Sigma^{-1/2}\|_2.
\end{aligned}$$

(f) Es gilt

$$\mathbb{E}|Z_\gamma| = \frac{2\gamma}{n}\mathbb{E}\|\mathbb{e}^T\mathbb{X}\Sigma^{-1/2}\|_2 \leq \frac{2\gamma}{n}\mathbb{E}[\|\mathbb{e}^T\mathbb{X}\Sigma^{-1/2}\|_2^2]^{1/2}$$

Hier ist

$$\mathbb{E}[\|\mathbb{e}^T\mathbb{X}\Sigma^{-1/2}\|_2^2] = \mathrm{tr}(\Sigma^{-1/2}\mathbb{X}^T \underbrace{\mathbb{E}[\mathbb{e}\mathbb{e}^T]}_{=\sigma^2 I_{d\times d}} \mathbb{X}\Sigma^{-1/2}) = n\sigma^2\mathrm{tr}(\Sigma^{-1/2}\Sigma\Sigma^{-1/2}) = nd\sigma^2.$$

Es folgt

$$\mathbb{E}|Z_\gamma| \leq \frac{2\gamma}{n}\cdot\sqrt{nd}\sigma = 2\gamma\sigma(\frac{d}{n})^{1/2}.$$

(g) Es soll gelten $\mathbb{P}(A^c) \leq \frac{1}{t}$. Markov-Ungleichung $\Rightarrow$

$$\mathbb{P}(A^c) \leq \frac{\mathbb{E}Z_\gamma}{a\gamma} \leq \frac{2\sigma(\frac{d}{n})^{1/2}}{a} = \frac{1}{t},$$

falls $a = 2\sigma(\frac{d}{n})^{1/2}t$ gewählt wird.
Es soll gelten: $(a\gamma)^{1/2} \leq \frac{\gamma}{2}$, d.h. $2a^{1/2} \leq \gamma^{1/2}$ bzw. $4a \leq \gamma$.
Wähle also $\gamma = 4a$.
Das ergibt die Rate

$$a\gamma = 4a^2 = 16\sigma^2\frac{d}{n}t^2.$$

(h) Es ist

$$\frac{1}{\sigma\sqrt{n}}\mathbb{e}^T\mathbb{X}\Sigma^{-1/2} = \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n \varepsilon_i\cdot(\Sigma^{-1/2}X_i) \sim N\Big(0, \underbrace{\frac{1}{n}\sum_{i=1}^n \Sigma^{-1/2}X_iX_i^T\Sigma^{-1/2}}_{=\Sigma^{-1/2}\underbrace{\frac{1}{n}\sum_{i=1}^n X_iX_i^T}_{=\Sigma}\Sigma^{-1/2}}\Big) = N(0, I_{d\times d}).$$

(i) Es gilt $\mathbb{E}[\|W\|_2^2] = \sum_{j=1}^d \mathbb{E}[W_j^2] = d$. Mit Markov-Ungleichung:

$$\mathbb{P}(\|W\|_2^2 - 2\mathbb{E}\|W\|_2^2 \geq t) = \mathbb{P}(\|W\|_2^2 \geq 2d + t) \leq \frac{\mathbb{E}[\exp(\frac{\|W\|_2^2}{4})]}{e^{-(2d+t)/4}}.$$

Unabhängigkeit der Komponenten von W liefert

$$\mathbb{E}[\exp(\frac{\|W\|_2^2}{4})] = \mathbb{E}[\exp(W_1^2/4)]^d = \sqrt{2}^d.$$

Es folgt

$$\mathbb{P}(\|W\|_2^2 - 2\mathbb{E}\|W\|_2^2 \geq t) \leq (\sqrt{2}e^{-1/2})^d \cdot e^{-t/4} \leq e^{-t/4}.$$

(j) Es ist

$$\begin{aligned} \mathbb{P}(Z_\gamma > a\gamma) &\overset{(e)}{\leq} \mathbb{P}(\frac{2}{n}\|\mathbb{e}^T \mathbb{X}\Sigma^{-1/2}\|_2 > a) = \mathbb{P}(\frac{\sigma}{\sqrt{n}}\|W\|_2 > a) \\ &= \mathbb{P}(\|W\|_2 > \sqrt{2d+t}) = \mathbb{P}(\|W\|_2^2 > 2d + t) \leq e^{-t/4}. \end{aligned}$$

(k) a wurde bereits gewählt, $\gamma = 4a$, Rate $\gamma a = 4a^2 = 4\frac{\sigma^2}{n}(2d+t)$.

Homepage der Vorlesung:
`https://ssp.math.uni-heidelberg.de/sam-ws2020/`