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

12. Übungsblatt

Aufgabe 1 (Maximalungleichungen aus der Bernstein-Ungleichung, Lemma 7.10(ii)).
Es seien $Z_1, ..., Z_n$ i.i.d. Zufallsvariablen mit Werten in $\mathcal{Z}$ und $\mathcal{G} \subset \{g : \mathcal{Z} \to \mathbb{R} \text{ messbar}\}$, so dass für alle $g \in \mathcal{G}$ gilt: $\|g\|_\infty \le M$. Definiere $H := \log(|\mathcal{G}| + 1)$, und

$$W := \sup_{g \in \mathcal{G}} \frac{|\sum_{i=1}^n \{g(Z_i) - \mathbb{E}g(Z_i)\}}{\mathbb{E}[g(Z_1)^2]^{1/2} + M\sqrt{\frac{H}{n}}}.$$

Wir wollen zeigen, dass $\mathbb{E}[W^2] \le 33nH$. Sei $t_0 > 0$.

(a) Zeigen Sie, dass

$$\mathbb{E}[W^2] \le t_0^2 + 2\int_{t_0}^\infty t\mathbb{P}(W \ge t)dt.$$

Hinweis: Es gilt $\mathbb{E}[W^2] = \int \mathbb{P}(W^2 \ge u)du$.

(b) Definiere $\tilde{g} := \frac{g}{\mathbb{E}[g(Z_1)^2]^{1/2} + M\sqrt{\frac{H}{n}}}$. Zeigen Sie mit Hilfe der Bernstein-Ungleichung für i.i.d. ZV $X_1, ..., X_n$

$$\mathbb{P}\Big(\big|\sum_{i=1}^n (X_i - \mathbb{E}X_1)\big| \ge x\Big) \le 2\exp\Big(-\frac{x^2}{2n\mathbb{E}[X_1^2] + \frac{2}{3}\|X_1\|_\infty \cdot x}\Big),$$

dass für $t \ge 3\sqrt{nH}$ gilt:

$$\mathbb{P}(W \ge t) \le 2|\mathcal{G}| \cdot \exp\Big(-\frac{3t}{4\sqrt{\frac{n}{H}}}\Big).$$

(c) Zeigen Sie, dass

$$\mathbb{E}[W^2] \le 33nH.$$

Hinweis: Definieren Sie $t_0 = 3\sqrt{nH}$ *und* $a := \frac{3}{4}\sqrt{\frac{H}{n}}$. *Mit partieller Integration gilt* $\int te^{-at}dt = -(a^{-2} + a^{-1}t)e^{-at}$.

Aufgabe 2 (Maximalungleichung für Normalverteilungen, Lemma 7.11).
Seien $Z_1, ..., Z_n \sim N(0, v^2)$ i.i.d. und $a_{ij} \in \mathbb{R}$, $i = 1, ..., n$, $j = 1, ..., N$. Sei

$$W_j := \frac{1}{v\sqrt{n}} \frac{\sum_{i=1}^n a_{ij}Z_i}{\big(\frac{1}{n}\sum_{i=1}^n a_{ij}^2\big)^{1/2}}.$$

Wir wollen zeigen, dass dann gilt:

$$\mathbb{E}\max_{j=1,...,N} |W_j| \le 2\sqrt{\log(N+1)}, \qquad \mathbb{E}\max_{j=1,...,N} |W_j|^2 \le 4\log(N+1).$$

(a) Zeigen Sie, dass $W_j \sim N(0,1)$.

(b) Sei $\varphi_2(x) = \exp(x^2) - 1$. Zeigen Sie $\varphi_2(\mathbb{E}\max_{j=1,...,N} \frac{|W_j|}{2}) \leq N$ und damit $\mathbb{E}\max_{j=1,...,N} \frac{|W_j|}{2} \leq 2\sqrt{\log(N+1)}$.
Hinweis: Für $W_j \sim N(0,1)$ gilt $\mathbb{E}\exp(\frac{W_j^2}{4}) = \sqrt{2}$.

(c) Sei $\varphi_1(x) = \exp(x) - 1$. Zeigen Sie $\varphi_1(\mathbb{E}\max_{j=1,...,N} \frac{|W_j|^2}{4}) \leq N$ und damit $\mathbb{E}\max_{j=1,...,N} |W_j|^2 \leq 4\log(N+1)$.

Aufgabe 3 (Covering Numbers von neuronalen Netzwerken, Lemma 7.12).
In dieser Aufgabe schätzen wir die Covering-Numbers $H(\gamma) = \log N(\gamma, \mathcal{F}(L,p,s,F), \|\cdot\|_\infty)$ der Klasse der neuronalen Netzwerke

$$\mathcal{F}(L,p,s,\infty) \quad := \quad \Big\{ f \in \mathcal{F}(L,p) : \sum_{l=0}^{L} \|W^{(l)}\|_0 + \sum_{l=1}^{L} \|v^{(l)}\|_0 \leq s, \qquad \forall l : \|W^{(l)}\|_\infty \leq 1, \|v^{(l)}\|_\infty \leq 1 \Big\}$$

ab, wobei

$$\begin{aligned} \mathcal{F}(L,p) \quad = \quad \Big\{ g : \mathbb{R}^{p_0} \to \mathbb{R}^{p_{L+1}} : g(x) = W^{(L)} \cdot \sigma_{v^{(L)}}\Big(W^{(L-1)} \cdot \sigma_{v^{(L-1)}}\Big(\dots W^{(1)} \cdot \sigma_{v^{(1)}}\big(W^{(0)} \cdot x\big) \dots \Big)\Big) \\ W^{(l)} \in \mathbb{R}^{p_l \times p_{l+1}} \quad (l = 0, ..., L), \qquad v^{(l)} \in \mathbb{R}^{p_l} \quad (l = 1, ..., L) \Big\} \end{aligned}$$

(a) Zeigen Sie, dass jedes $\mathcal{F}(L,p)$ geschrieben werden kann als $\{f_\theta : \theta \in \Theta^{pre}\}$ mit $\Theta^{pre} \subset [-1,1]^T$, wobei

$$T = \sum_{l=0}^{L} p_l p_{l+1} + \sum_{l=1}^{L} p_l.$$

(b) Zeigen Sie: Es gilt $\mathcal{F}(L,p,s,\infty) = \{f_\theta : \theta \in \Theta\}$, wobei

$$\Theta \subset \bigcup_{S \subset \{1,...,T\} : |S| \leq s} \Theta_S, \qquad \Theta_S = \{\theta \in [-1,1]^T : \forall j \in S^c : \theta_j = 0\}.$$

(c) Zeigen Sie: Für $a > 0$ und $S \subset \{1,...,T\}$, $|S| \leq s$ gibt es $\tilde{\Theta}_S \subset [-1,1]^T$ mit

$$\forall \theta \in \Theta_S \quad \exists \tilde{\theta} \in \tilde{\Theta}_S : \quad \|\theta - \tilde{\theta}\|_\infty \leq a, \qquad |\tilde{\Theta}_S| \leq \left\lfloor \frac{2}{a} \right\rfloor^s.$$

Definiere $\tilde{\Theta} := \bigcup_{S \subset \{1,...,T\} : |S| \leq s} \tilde{\Theta}_S$. Dann gilt: $\forall \theta \in \Theta \quad \exists \tilde{\theta} \in \tilde{\Theta} : \quad \|\theta - \tilde{\theta}\|_\infty \leq a$. Zeigen Sie weiter:

(d) $T \leq V := \prod_{l=0}^{L+1}(p_l + 1)$ und

$$|\tilde{\Theta}| \leq \Big(\frac{2V}{a}\Big)^{s+1}.$$

Hinweis: Es gilt $\binom{T}{k} \leq T^k$.

Für $f_\theta \in \mathcal{F}(L,p,s,F)$ wähle $\tilde{\theta} \in \tilde{\Theta}$ mit $\|\theta - \tilde{\theta}\|_\infty \leq a$. Wir wollen zeigen: $\|f_\theta - f_{\tilde{\theta}}\|_\infty \leq a \cdot (L+1) \cdot V$. (*)

(e) Zeigen Sie, dass aus (*) und (d) folgt:

$$N(\gamma, \mathcal{F}(L,p,s,F), \|\cdot\|_\infty) \leq (2\gamma^{-1}V^2(L+1))^{s+1}.$$

Definiere für $\theta = (v^1, ..., v^{(L)}, W^{(0)}, ..., W^{(L)})$:

$$A_\theta^{k+}(x) = \sigma_{v^{(k)}} W^{(k-1)} \sigma_{v^{(k-1)}} ... W^{(1)} \sigma_{v^{(1)}} W^{(0)} x$$

und

$$A_\theta^{k-}(y) = W^{(L)} \sigma_{v^{(L)}} ... W^{(k)} \sigma_{v^{(k)}} W^{(k-1)} x.$$

(f) Zeigen Sie, dass $|A_\theta^{k-}(x) - A_\theta^{k-}(x')| \leq \big(\prod_{l=k-1}^{L} p_l\big) \cdot \|x - x'\|_\infty$.
Hinweis: Die Zeilensummennorm $\|A\|_Z$ einer Matrix A erfüllt $\|Ax\|_\infty \leq \|A\|_Z \|x\|_\infty$.

(g) Zeigen Sie: Es gilt $|A_\theta^{k+}(x)|_\infty \leq \prod_{l=0}^{k-1}(p_l + 1)$.

(h) Folgern Sie aus (f),(g):

$$|f_\theta(x) - f_{\tilde{\theta}}(x)| \leq a \cdot (L+1) \cdot V.$$

(i) Zeigen Sie, dass gilt

$$N(\gamma, \mathcal{F}(L,p,s,\infty), \|\cdot\|_\infty) = N(\gamma, \mathcal{F}(L, (p_0, p_1 \wedge s, ..., p_L \wedge s, p_{L+1}), s, \infty), \|\cdot\|_\infty)$$

(j) Zeigen Sie: Für $s \geq 2$, $L \geq 1$ gibt es eine universelle Konstante $c > 0$, so dass $H(\gamma) = \log N(\gamma, \mathcal{F}(L,p,s,\infty), \|\cdot\|_\infty)$ erfüllt:

$$H(\gamma) \leq c \cdot s \cdot \big\{L \log(s) + \log(\gamma^{-1}) + \log(p_0 p_{L+1})\big\}.$$

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