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

10. Übungsblatt

Aufgabe 1 (Covering Numbers von Baumstümpfen).
Sei $\mathcal{X} \subset [0,1]^d$ und $\mathcal{Y} = \{-1,+1\}$. In dieser Aufgabe wollen wir zeigen, dass die Klasse

$$\begin{aligned} \mathcal{C} &:= \{f_T : \ T \text{ CART mit 1 inneren Knoten } \} \\ &= \{y_1 \mathbb{1}_{\{x_j < s\}} + y_2 \mathbb{1}_{\{x_j \geq s\}} : y_1, y_2 \in \mathcal{Y}, s \in \mathbb{R}, j \in \{1, ..., d\}\} \end{aligned}$$

erfüllt:

$$N(\varepsilon, \mathcal{C}, \|\cdot\|_{2,n,X}) \leq (\frac{c_{cov}}{\varepsilon})^V, \qquad \|g\|_{2,n,X} := \Big(\frac{1}{n}\sum_{i=1}^n g(X_i)^2\Big)^{1/2} \qquad (*)$$

wobei $V = 2 \cdot \lfloor 2\log_2(2d) \rfloor$ und c_{cov} eine universelle Konstante. Dazu nutzen wir folgendes Resultat: Ist $\mathcal{V}(\mathcal{C})$ die VC-Dimension von $\mathcal{C}$, so gilt

$$N(\varepsilon, \mathcal{C}, \|\cdot\|_{2,n,X}) \leq 13 \cdot \mathcal{V}(\mathcal{C}) \cdot (\frac{4e}{\varepsilon^2})^{\mathcal{V}(\mathcal{C})}.$$

Hierbei ist $\mathcal{V}(\mathcal{C})$ wie folgt definiert:

$$\mathcal{V}(\mathcal{C}) = \inf\{N \in \mathbb{N} : m_{\mathcal{C}}(N) < 2^N\},$$

wobei

$$m_{\mathcal{C}}(N) := \max_{x_1,...,x_N \in [0,1]^d} m_{\mathcal{C}}(x_1, ..., x_N), \qquad m_{\mathcal{C}}(x_1, ..., x_N) := \#\{(f(x_1), ..., f(x_N)) : f \in \mathcal{C}\}.$$

die maximale Anzahl an unterschiedlichen Labelings von N Punkten im Raum, die durch $\mathcal{C}$ erreicht werden kann. Lösen Sie dazu die folgenden Aufgaben.

(a) Zeigen Sie: Falls $\mathcal{V}(\mathcal{C}) \leq \lfloor 2\log_2(2d) \rfloor$, so folgt (*)

(b) Sei zunächst $d = 1$. Zeigen Sie, dass

$$m_{\mathcal{C}}(N) = 2N.$$

(c) Sei d beliebig. Zeigen Sie, dass

$$m_{\mathcal{C}}(N) \leq \min\{2Nd, 2^N\}.$$

(d) Zeigen Sie, dass für $d \geq 1$ gilt: $\mathcal{V}(\mathcal{C}) \leq \lfloor 2\log_2(2d) \rfloor$.
Hinweis: $d - 1 \ \geq \log_2(d)$ *für alle* $d \geq 1$.

Aufgabe 2 (Beweis der Margin Condition, Lemma 6.27(ii)).
Sei $\phi \in \{\phi_1(x) = e^x, \phi_2(x) = \log(1+e^x)\}$ monoton wachsend, stetig differenzierbar und konvex. Sei $\tilde{L}(y,s) = \phi(-ys)$. Wir zeigen, dass für alle $\delta : \mathcal{X} \to \mathbb{R}$, $\|\delta\|_\infty \leq \rho$ gilt.

$$\mathbb{E}[(\tilde{L}(Y,\delta(X)) - \tilde{L}(Y,\delta^*(X)))^2] \leq c_\rho \cdot \{\tilde{R}(\delta) - \tilde{R}(\delta^*)\},$$

wobei $\delta^* \in \operatorname{argmin}_{\delta:\mathcal{X}\to\mathbb{R}} \tilde{R}(\delta)$, $\tilde{R}(\delta) = \mathbb{E}\tilde{L}(Y,\delta(X))$ und $c_\rho = \phi(\rho) + \phi(-\rho) + C_\phi$, mit

$$C_\phi = \begin{cases} 0, & \phi = \phi_1 \\ 2 - 2\log(2), & \phi = \phi_2 \end{cases}.$$

Dazu gehen wir wie folgt vor:

(a) Zeigen Sie, dass $\delta^*(x) = g(\eta(x))$, wobei $g(\eta)$ erfüllt (vgl. Risiko-Übertragung Satz 3.21):

$$0 = \phi'(-g(\eta)) \cdot \eta + \phi'(g(\eta)) \cdot (1-\eta).$$

(b) Zeigen Sie, dass

$$\begin{aligned} \mathbb{E}[(\tilde{L}(Y,\delta(X)) - \tilde{L}(Y,\delta^*(X)))^2 | X = x] &= A_2(\eta(x),\delta(x)), \\ \mathbb{E}[\tilde{L}(Y,\delta(X)) - \tilde{L}(Y,\delta^*(X)) | X = x] &= A_1(\eta(x),\delta(x)), \end{aligned}$$

und bestimmen Sie die Funktionen A_1, A_2.
Hinweis: Es gilt zum Beispiel $A_1(\eta,\delta) = \eta(\phi(-\delta) - \phi(-g(\eta))) + (1-\eta)(\phi(\delta) - \phi(g(\eta)))$.

(c) Zeigen Sie mit (a), dass

$$\partial_\eta A_1(\eta,\delta) = \big[\phi(-\delta) - \phi(-g(\eta))\big] - \big[\phi(\delta) - \phi(g(\eta))\big].$$

(d) *Achtung hässlich!* Zeigen Sie mit (a), dass

$$\partial_\eta A_2(\eta,\delta) = \partial_\eta A_1(\eta,\delta) \cdot (\phi(\delta) + \phi(-\delta) + B(\eta)),$$

wobei

$$B(\eta) := \big\{\{\eta\phi'(-g(\eta)) + (1-\eta)\phi'(g(\eta))\}g'(\eta) - \phi(g(\eta)) - \phi(-g(\eta))\big\}$$

(e) Setze $C_\phi := 0 \vee \max_{\eta\in[0,1]} B(\eta)$. Zeigen Sie: Falls $g(\eta) \geq \delta$, gilt

$$\partial_\eta A_1(\eta,\delta) \leq \{\phi(\rho) + \phi(-\rho) + C_\phi\} \cdot \partial_\eta A_2(\eta,\delta).$$

Ansonsten gilt die umgekehrte Ungleichung.
Hinweis: Es gilt $\phi(\delta) + \phi(-\delta) \leq \phi(\rho) + \phi(-\rho)$. *Warum?*

(f) Es bezeichne g^{-1} die Umkehrfunktion von g. Zeigen Sie, dass $A_j(g^{-1}(\delta),\delta) = 0$, $j = 1,2$.

(g) Zeigen Sie, dass für $g(\eta) \geq \delta$ gilt:

$$A_2(\eta,\delta) \leq \{\phi(\rho) + \phi(-\rho) + C_\phi\} \cdot A_1(\eta,\delta).$$

Analoges gilt für die umgekehrte Ungleichung. *Hinweis: Integration.*

(h) Weisen Sie nach, dass die oben angegebenen Werte für C_ϕ korrekt sind.

Aufgabe 3 (Beweis Lemma 6.31 / Covering Numbers übertragen).
Es sei $\tilde{L}(y, s) = \phi(-ys)$ mit konvexem, monoton wachsendem, stetig differenzierbaren ϕ. Es sei

$$\mathcal{F} = \{\tilde{L}(y, \delta(x)) - \tilde{L}(y, \delta_0(x)) : \delta \in B(\rho), D(\delta, \delta_0) \leq r\}$$

die Funktionenklasse aus Lemma 6.31 (wir lassen im Folgenden $D(...) \leq r$ weg). Sei $\|f\|_{2,n} = \left(\frac{1}{n}\sum_{i=1}^n f(X_i, Y_i)^2\right)^{1/2}$. Definiere

- $\mathcal{F}_1 = \{\tilde{L}(y, \delta(x)) = \phi(-y\delta(x)) : \delta \in B(\rho)\}$.
- $\mathcal{F}_2 = \{-y\delta(x) : \delta \in B(\rho)\}$
- $\mathcal{F}_3 = \{\delta(x) : \delta \in B(\rho)\}$ mit neuem Abstandsmaß $\|f\|_{2,n,X} := \left(\frac{1}{n}\sum_{i=1}^n f(X_i)^2\right)^{1/2}$
- $\mathcal{F}_4 = \{\frac{\delta(x)}{\rho} : \delta \in B(\rho)\}$.

Zeigen Sie:

(a) $N(\varepsilon, \mathcal{F}, \|\cdot\|_{2,n}) \leq N(\varepsilon, \mathcal{F}_1, \|\cdot\|_{2,n})$.

(b) $N(\varepsilon, \mathcal{F}_1, \|\cdot\|_{2,n}) \leq N(\frac{\varepsilon}{\phi'(\rho)}, \mathcal{F}_2, \|\cdot\|_{2,n})$.

(c) $N(\frac{\varepsilon}{\phi'(\rho)}, \mathcal{F}_2, \|\cdot\|_{2,n}) \leq N(\frac{\varepsilon}{\phi'(\rho)}, \mathcal{F}_3, \|\cdot\|_{2,n,X})$, wobei $\|f\|_{2,n,X} := (\frac{1}{n}\sum_{i=1}^n f(X_i)^2)^{1/2}$.

(d) $N(\frac{\varepsilon}{\phi'(\rho)}, \mathcal{F}_3, \|\cdot\|_{2,n,X}) \leq N(\frac{\varepsilon}{\rho\phi'(\rho)}, \mathcal{F}_4, \|\cdot\|_{2,n,X})$.

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