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

11. Übungsblatt

Aufgabe 1 (Approximation von Funktionen beschränkter Variation)
Lösung:

(a) u ist monoton wachsend. Daher gilt für alle $M \in \mathbb{N}$, $0 \le s_0 \le ... \le s_M \le 1$,

$$\sum_{i=1}^{M} \big|u(s_i) - u(s_{i-1})\big| = \sum_{i=1}^{M} \big(u(s_i) - u(s_{i-1})\big) = u(s_M) - u(s_0) \le u(1) - u(0),$$

mit Gleichheit für $s_M = 1$, $s_0 = 0$. $\Rightarrow |u|_{BV} = u(1) - u(0)$.

(b) Für $z \in [t_j, t_{j+1})$ gilt wegen $t_1 \le t_2 \le ... \le t_N$:

$$\tilde{u}(z) = u(0) + \frac{C}{N} \cdot j,$$

und nach Definition von t_j

$$C\frac{j}{N} < u(z) - u(0) \le C\frac{j+1}{N} \quad \Rightarrow \quad u(0) + C\frac{j+1}{N} < u(z) \le u(0) + C\frac{j+1}{N}$$

(Achtung: Beide Aussagen gelten nie, falls $t_j = t_{j+1}$. Dies komprimittiert aber nicht den Beweis).
Es folgt

$$|u(z) - \tilde{u}(z)| \le \frac{C}{N}.$$

Für jedes $z \in [0,1)$ gibt es $j \in \{0, ..., N-1\}$ mit $z \in [t_j, t_{j+1}) \Rightarrow$

$$\|u - \tilde{u}\|_\infty \le \frac{C}{N} = \frac{|u|_{BV}}{N}.$$

(c) Sei $\tilde{\tilde{u}}$ die in der Aufgabenstellung angegebene alternative Darstellung. Für $z \in [t_j, t_{j+1})$ gilt

$$\begin{aligned}
\tilde{\tilde{u}}(z) &= \frac{u(0) + u(1)}{2} + \sum_{i=1}^{j} \frac{C}{2N} - \sum_{i=j+1}^{N} \frac{C}{2N} \\
&= \frac{u(0) + u(1)}{2} + \frac{C}{2N}(j - (N - j)) \\
&= \frac{u(0) + u(1)}{2} - \frac{C}{2} + \frac{C}{N}j \\
&= u(0) + \frac{C}{N}j = \tilde{u}(z).
\end{aligned}$$

(d) Wir führen dasselbe Konstruktionsprinzip für v durch mit 'Quantilen' $q_i := \sup\{z \in [0,1] : v(z) - v(0) \leq \frac{|v|_{BV}}{N} i\}$, $i = 0, ..., N$ und erhalten

$$\tilde{v}(z) = \frac{v(0) + v(1)}{2}(\mathbb{1}_{\{z \geq 0\}} - \mathbb{1}_{\{z<0\}}) + \sum_{i=1}^{N} \frac{|v|_{BV}}{2N}(\mathbb{1}_{\{z \geq q_i\}} - \mathbb{1}_{\{z<q_i\}}).$$

Definiere $\tilde{g} := \tilde{u} - \tilde{v}$, dann gilt

$$\|\tilde{g} - g\|_\infty \leq \|\tilde{u} - u\|_\infty + \|\tilde{v} - v\|_\infty \leq \frac{|u|_{BV}}{N} + \frac{|v|_{BV}}{N} = \frac{|g|_{BV}}{N},$$

und

$$\begin{aligned}\tilde{g}(z) &= \tilde{u}(z) - \tilde{v}(z) = \underbrace{\Big[\frac{u(0) + u(1)}{2} - \frac{v(0) + v(1)}{2}\Big]}_{=\frac{g(0)+g(1)}{2}}(\mathbb{1}_{\{z \geq 0\}} - \mathbb{1}_{\{z<0\}}) \\ &\quad + \sum_{i=1}^{N} \frac{|u|_{BV}}{2N}(\mathbb{1}_{\{z \geq t_i\}} - \mathbb{1}_{\{z<t_i\}}) - \sum_{i=1}^{N} \frac{|v|_{BV}}{2N}(\mathbb{1}_{\{z \geq q_i\}} - \mathbb{1}_{\{z<q_i\}}).\end{aligned}$$

Aufgabe 2 (Konvergenzrate beim Boosting von Baumstümpfen)
Lösung:

(a) Aufgabe 1(e) $\Rightarrow$ Es gibt monoton wachsende Funktionen u_j, v_j mit $h_j = u_j - v_j$ und $|h_j|_{BV} = |u_j|_{BV} + |v_j|_{BV}$ und zugehörige Funktionen $\tilde{h}_j : [0,1] \to \mathbb{R}$ mit

$$\|h_j - \tilde{h}_j\|_\infty \leq \frac{B}{N},$$

und

$$\|\tilde{h}_j\|_1 \leq \frac{|h_j(0)| + |h_j(1)|}{2} + \sum_{i=1}^{N} \big(\frac{|u|_{BV}}{2N} + \frac{|v|}{2N}\big) = \frac{|h_j(0)| + |h_j(1)| + |h_j|_{BV}}{2} \leq \frac{B}{2}.$$

(b) Es gilt (beachte hier den Dimensionssprung von d zu 1 in den Bezeichnungen $\|\cdot\|_\infty$ und $\|\cdot\|_1$):

$$\|\delta^* - \tilde{h}\|_\infty \leq \sum_{j=1}^{d} \|h_j - \tilde{h}_j\|_\infty \leq \frac{Bd}{N},$$

und

$$\|\tilde{h}\|_1 \leq \sum_{j=1}^{d} \|\tilde{h}_j\|_1 \leq \frac{Bd}{2}.$$

(c) Zu δ^* wähle die Funktion $\tilde{h} \in \Delta$. Dann gilt

$$\inf_{\delta \in \Delta} \big\{\tilde{R}(\delta) - \tilde{R}(\delta^*) + 2\lambda P(\|\delta\|_1)\big\} \leq \tilde{R}(\tilde{h}_N) - \tilde{R}(\delta^*) + 2\lambda P(\|\tilde{h}_N\|_1).$$

Da $\|\delta^*\|_\infty \leq \sum_{j=1}^{d} \|h_j\|_\infty \leq \frac{Bd}{2}$ und $\|\tilde{h}_N\|_\infty \leq \|\tilde{h}_N - \delta^*\|_\infty + \|\delta^*\|_\infty \leq \frac{Bd}{N} + \frac{Bd}{2}$, gilt auch

$$\phi(-y\tilde{h}_N(x)) \leq \phi(\|\tilde{h}_N\|_\infty) < \infty, \qquad \phi(-y\delta^*(x)) \leq \phi(\|\delta^*\|_\infty) < \infty.$$

Dominierte Konvergenz, ϕ L-stetig, $\tilde{h}_N \to \delta^*$ gleichmäßig $\Rightarrow$

$$\tilde{R}(\tilde{h}_N) - \tilde{R}(\delta^*) = \mathbb{E}\big[\underbrace{\phi(-Y\tilde{h}_N(X)) - \phi(-Y\delta^*(X))}_{\to 0 \quad (N\to\infty)}\big] \to 0.$$

$\Rightarrow$

$$\begin{aligned}&\inf_{\delta\in\Delta}\big\{\tilde{R}(\delta) - \tilde{R}(\delta^*) + 2\lambda P(\|\delta\|_1)\big\}\\ \leq\ &\limsup_{N\to\infty}\big\{\tilde{R}(\tilde{h}_N) - \tilde{R}(\delta^*) + 2\lambda P(\|\tilde{h}_N\|_1)\big\} \leq 2\lambda P(\|\tilde{h}_N\|_1) \leq 2\lambda P(\frac{Bd}{2}).\end{aligned}$$

(d) **Betrachtung von λ:** Es gilt

$$V = 2\lfloor 2\log_2(2d)\rfloor \leq 4\log_2(2d) = \frac{4}{\log(2)}\log(2d) \leq \frac{4}{\log(2)}(\log(2){+}\log(d{+}1)) \leq \frac{8}{\log(2)}\log(d{+}1).$$

Mit $1 \leq \frac{V+2}{V+1} \leq 2$ folgt:

$$((V+2)V^{1/2})^{\frac{V+2}{V+1}} \leq (3V^{3/2})^{\frac{V+2}{V+1}} \leq 9V^3 \leq 9\big(\frac{8}{\log(2)}\big)^3\log(d+1)^3, \qquad n^{-\frac{1}{2}\frac{V+2}{V+1}} \leq n^{-\frac{1}{2}}.$$

Einsetzen in λ liefert

$$\lambda \leq c\cdot 9\big(\frac{8}{\log(2)}\big)^3\cdot(1+t)\log(d+1)^3\cdot n^{-\frac{1}{2}}$$

Betrachtung vom Bestrafungsterm: Falls $\phi(x) = e^x$, gilt

$$P(\frac{\rho}{2}) = (\rho\phi'(\rho))^{\frac{V}{V+1}}\phi(\rho)^{\frac{1}{V+1}} + \phi(\rho) \leq 2(\rho+1)e^\rho$$

$\Rightarrow P(\frac{Bd}{2}) \leq 2(Bd+1)e^{Bd}$.
Falls $\phi(x) = \log(1+e^x)$, gilt $\phi'(x) = \frac{e^x}{1+e^x} \leq 1$ und $\phi(x) \leq 1 + x$, daher

$$P(\frac{\rho}{2}) = (\rho\phi'(\rho))^{\frac{V}{V+1}}\phi(\rho)^{\frac{1}{V+1}} + \phi(\rho) \leq 2(\rho+1).$$

$\Rightarrow P(\frac{Bd}{2}) \leq 2(Bd+1)$. Einsetzen in (c) $2\lambda P(\frac{Bd}{2})$ liefert die Behauptung.

(e) In (b) erhält man nun

$$\|h - \tilde{h}\|_\infty \leq \frac{Bs}{N}, \qquad \|\tilde{h}\|_1 \leq \frac{Bs}{2}.$$

$\Rightarrow$ Nur der Term in der Fallunterscheidung von (d) ändert sich, d wird durch s ersetzt:

$$\inf_{\delta\in\Delta}\big\{\tilde{R}(\delta) - \tilde{R}(\delta^*) + 2\lambda P(\|\delta\|_1)\big\} \leq c'(1+t)\log(d+1)^3\cdot n^{-\frac{1}{2}}\cdot\begin{cases}(Bs+1)e^{Bs+1}, & \phi = \phi_1,\\ (Bs+1), & \phi = \phi_2.\end{cases}$$

Man bezahlt also mit $\log(d+1)^3$ für die zugrundeliegende Dimension d (dies entsteht aus der Orakel-Ungleichung, d.h. letztlich dem Schätzfehler), und nun nur mit s (statt vorher d) für die kleinere Anzahl der Nicht-Null-Summanden von δ^*.

Aufgabe 3 (Approximation mit neuronalen Netzwerken)
Lösung:

(a) Skizze.

(b) Das Netzwerk hat $1 + (m-1) = m$ hidden layer, mit der jeweils angegebenen Größe in der Skizze.

(c) Es gilt

$$\begin{aligned}
& g(\frac{x-y+1}{2}) - g(\frac{x+y}{2}) + \frac{x+y}{2} - \frac{1}{4} \\
= \; & \frac{1+(x-y)}{2} \cdot \frac{1-(x-y)}{2} - \frac{x+y}{2} \cdot (1 - \frac{x+y}{2}) + \frac{x+y}{2} - \frac{1}{4} \\
= \; & \frac{1}{4}(1-(x-y)^2) + \frac{1}{4}(x+y)^2 - \frac{1}{4} \\
= \; & \frac{1}{4}((x+y)^2 - (x-y)^2) = \frac{1}{4} \cdot 4xy = xy.
\end{aligned}$$

Dadurch kann die Approximation der Funktion $x \cdot y$ mit Netzwerken auf die Approximation der Funktion g mit Netzwerken zurückgeführt werden.

(d) Es gibt $q = \lceil \log_2(r) \rceil$ Schritte ($2^q \to 2^{q-1}$, ..., $2^1 \to 2^0$). In jedem dieser Schritte werden Multiplikationsnetzwerke $f_m \in \mathcal{F}((m+4), (2, 6, 6, ..., 6, 1))$ eingefügt (Erhöhe wegen Verbindung Anzahl Layer um 1 und multipliziere dann mit q). Im ersten Schritt werden dafür maximal r Netzwerke nebeneinander gebraucht (multipliziere Breite der Hidden layer mit r). Daher ergibt sich ein

$$\mathcal{F}(q(m+5), (r, 6r, 6r, ..., 6r, 1))$$

Netzwerk.
Approximation: Der Einfachheit halber sei $r = 2^q$ mit $q \in \mathbb{N}$. Dann gilt mit $y_1 = (x_1, ..., x_{r/2}), y_2 = (x_{r/2+1}, ..., x_r)$ und wegen $0 \leq f_{m,r/2}(y_1) \leq 1$:

$$\begin{aligned}
\left|f_{m,r}(x) - \prod_{j=1}^{r} x_j\right| &= \left|f_m(f_{m,r/2}(y_1), f_{m,r/2}(y_2)) - \prod_{j=1}^{r/2} x_j \cdot \prod_{j=r/2+1}^{r} x_j\right| \\
&\leq \left|f_m(f_{m,r/2}(y_1), f_{m,r/2}(y_2)) - f_{m,r/2}(y_1) \cdot f_{m,r/2}(y_2)\right| \\
&\qquad + \left|f_{m,r/2}(y_1) \cdot f_{m,r/2}(y_2) - \prod_{j=1}^{r/2} x_j \cdot \prod_{j=r/2+1}^{r} x_j\right| \\
&\leq 2^{-m} + \left|f_{m,r/2}(y_1) - \prod_{j=1}^{r/2} x_j\right| + \left|f_{m,r/2}(y_2) - \prod_{j=r/2+1}^{r} x_j\right|
\end{aligned}$$

Wegen $|f_{m,2}(x) - x_1 x_2| \leq 2^{-m}$ tauchen die zwei rekursiven Terme im Schritt $r/2 = 1$ nicht mehr auf.

Induktiv: $\left|f_{m,r}(x) - \prod_{j=1}^{r} x_j\right| \leq 3^q 2^{-m} \leq r^2 2^{-m}$.

(e) Taylor-Entwicklung $\Rightarrow$ Es gibt $\xi_x \in [0,1]$ mit

$$f(x) = T_a(x) + \sum_{\alpha \in \mathbb{N}_0^r : |\alpha| = \beta} (\partial^\alpha f)(a + \xi(x-a)) \cdot \frac{(x-a)^\alpha}{\alpha!}.$$

$\Rightarrow$

$$|f(x) - T_a(x)| \leq \sum_{\alpha \in \mathbb{N}_0^r : |\alpha| = \beta} \underbrace{\frac{|(x-a)^\alpha|}{\alpha!}}_{\leq \frac{|x-a|_\infty^\beta}{\alpha!}} \cdot \underbrace{\big|(\partial^\alpha f)(a + \xi(x-a))\big|}_{\leq K} \leq K|x-a|_\infty^\beta \cdot \underbrace{\sum_{|\alpha|=\beta} \frac{1}{\alpha!}}_{\leq e^r}.$$

(f) Es gilt

$$\sum_{a \in D(M)} \prod_{j=1}^r \big(1 - M \cdot |x_j - a_j|\big)_+ = \prod_{j=1}^r \sum_{l=0}^M (1 - M|x_j - \ell/M|)_+ = 1,$$

daher

$$|T(x) - f(x)| \leq \sum_{a \in D(M), \|a-x\|_\infty \leq 1/M} \underbrace{\big|T_a(x) - f(x)\big|}_{\overset{(e)}{\leq} Ke^r M^{-\beta}} \cdot \prod_{j=1}^r \big(1 - M \cdot |x_j - a_j|\big)_+ \leq Ke^r M^{-\beta}.$$

(g) Aussage (f) reduziert die Approximation von beliebigen stetig differenzierbaren Funktionen auf die Approximation von Produkten (bzw. Produkten von Beträgen). Aussage (d) liefert, wie Produkte durch neuronale Netzwerke approximiert werden können.

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