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

7. Übungsblatt

Aufgabe 1 (Alternativer Beweis des vereinfachten SVM-Schätzers)
Lösung:

(a) $\mathcal{H}_K$ Vektorraum und $\delta_0, \hat{\delta} \in \mathcal{H}_K \Rightarrow \tilde{\delta} = T\hat{\delta} + (1-T)\delta_0 \in \mathcal{H}_K$. Außerdem

$$\|\tilde{\delta}\|_K \leq T\|\hat{\delta}\|_K + (1-T)\|\delta_0\|_K \leq T\rho + (1-T)\rho = \rho.$$

$\Rightarrow \tilde{\delta} \in B(\rho)$.
$\tilde{\delta} - \delta_0 = T(\hat{\delta} - \delta_0) \Rightarrow$

$$D(\tilde{\delta}, \delta_0) = TD(\hat{\delta}, \delta_0) = \frac{rD(\hat{\delta}, \delta_0)}{r + D(\hat{\delta}, \delta_0)} \leq r.$$

(b) Es gilt $\tilde{R}_n(\hat{\delta}) \leq \tilde{R}_n(\delta_0) \Rightarrow$

$$\tilde{R}_n(\tilde{\delta}) = \tilde{R}_n(T\hat{\delta} + (1-T)\delta_0) \overset{s\mapsto\tilde{L}(y,s) \text{ konvex}}{\leq} T \underbrace{\tilde{R}_n(\hat{\delta})}_{\leq \tilde{R}_n(\delta_0)} + (1-T)\tilde{R}_n(\delta_0) \leq \tilde{R}_n(\delta_0).$$

Es folgt

$$\tilde{R}(\tilde{\delta}) - \tilde{R}(\delta_0) \quad = \quad \underbrace{\{\tilde{R}_n(\tilde{\delta}) - \tilde{R}_n(\delta_0)\}}_{\leq 0} + \underbrace{\{\tilde{R}(\tilde{\delta}) - \tilde{R}_n(\tilde{\delta}) - (\tilde{R}(\delta_0) - \tilde{R}_n(\delta_0))\}}_{\overset{(a)}{\leq} Z_r}$$

$\Rightarrow$

$$\tilde{R}(\tilde{\delta}) - \tilde{R}(\delta^*) = \{\tilde{R}(\tilde{\delta}) - \tilde{R}(\delta_0)\} + \{\tilde{R}(\delta_0) - \tilde{R}(\delta^*)\} \leq \{\tilde{R}(\delta_0) - \tilde{R}(\delta^*)\} + Z_r.$$

(c) Auf A und unter der Annahme gilt

$$\tilde{R}(\tilde{\delta}) - \tilde{R}(\delta^*) \leq \frac{1}{8c_\rho}(\frac{r}{2})^2 + \frac{1}{8c_\rho}(\frac{r}{2})^2 = \frac{1}{4c_\rho}(\frac{r}{2})^2.$$

Quadratic margin condition $\Rightarrow$

$$\frac{1}{c_\rho}D(\tilde{\delta}, \delta^*)^2 \leq \tilde{R}(\tilde{\delta}) - \tilde{R}(\delta^*) \leq \frac{1}{4c_\rho}(\frac{r}{2})^2.$$

Umstellen $\Rightarrow$

$$D(\tilde{\delta}, \delta^*) \leq \frac{r}{4}.$$

(d) Zunächst gilt (quadratic margin condition):

$$\frac{1}{c_\rho}D(\delta_0,\delta^*)^2 \le \tilde{R}(\delta_0)-\tilde{R}(\delta^*) \le \frac{1}{4c_\rho}(\frac{r}{2})^2.$$

$\Rightarrow D(\delta_0,\delta^*) \le \frac{r}{4}$.
$\Rightarrow$

$$D(\tilde{\delta},\delta_0) \le D(\tilde{\delta},\delta^*) + D(\delta_0,\delta^*) \le \frac{r}{4}+\frac{r}{4} = \frac{r}{2}.$$

Es ist $\tilde{\delta}-\delta_0 = T(\hat{\delta}-\delta_0)$, daher

$$\frac{r}{2} \ge D(\tilde{\delta},\delta^*) = TD(\hat{\delta},\delta_0) = \frac{rD(\hat{\delta},\delta_0)}{r+D(\hat{\delta},\delta_0)}$$

Umstellen $\Rightarrow D(\hat{\delta},\delta_0) \le r$.
Durchführung (b) analog für $\hat{\delta}$ liefert auf A:

$$\tilde{R}(\hat{\delta})-\tilde{R}(\delta^*) \le \{\tilde{R}(\delta_0)-\tilde{R}(\delta^*)\} + \frac{1}{8c_\rho}(\frac{r}{2})^2.$$

(e) Es gilt

$$\sup_{\delta\in B(\rho), D(\delta,\delta_0)\le r}\{\tilde{R}(\delta)-\tilde{R}_n(\delta)-(\tilde{R}(\delta_0)-\tilde{R}_n(\delta_0))\} = \sup_{\delta\in\mathcal{F}} f_\delta(X_i,Y_i),$$

wobei $\mathcal{F} = \{\delta\in B(\rho) : D(\delta,\delta_0)\le r\}$ und

$$f_\delta(x,y) = \frac{1}{n}\Big\{\mathbb{E}\tilde{L}(Y,\delta(X)) - \tilde{L}(y,\delta(x)) - (\mathbb{E}\tilde{L}(Y,\delta_0(X)) - \tilde{L}(y,\delta_0(x)))\Big\}.$$

Beachte: Für $\delta\in B(\rho)$ gilt:

$$|\tilde{L}(y,\delta(x))| \le 1+|\delta(x)| \le 1+\|\delta\|_\infty \le 1+\rho.$$

Damit

$$\|f_\delta\|_\infty \le \frac{4(1+\rho)}{n},$$

und

$$\mathrm{Var}(f_\delta(X,Y)) \le \frac{1}{n^2}\mathbb{E}[(\tilde{L}(Y,\delta(X))-\tilde{L}(Y,\delta_0(X)))^2] \le \frac{1}{n^2}\mathbb{E}[(\delta(X)-\delta_0(X))^2] = \frac{D(\delta,\delta_0)^2}{n^2} \le \frac{r^2}{n^2}.$$

Talagrand, $\alpha>0 \Rightarrow$ Mit WS $\ge 1-e^{-t}$,

$$\begin{aligned} Z_r &\le (1+\alpha)\underbrace{\mathbb{E}Z_r}_{\le\phi_\rho(r^2)} + \sqrt{2tn}\cdot\frac{r}{n} + 4(\frac{1}{\alpha}+\frac{1}{3})(\rho+1)\cdot\frac{t}{n} \\ &\overset{\alpha=1}{\le} 2\phi_\rho(r^2) + \sqrt{\frac{2t}{n}}\cdot r + \frac{16(\rho+1)}{3}\cdot\frac{t}{n}. \end{aligned}$$

(f) r muss erfüllen:

$$2\phi_\rho(r^2) + \sqrt{\frac{2t}{n}}\cdot r + \frac{16(\rho+1)}{3}\cdot\frac{t}{n} \le \frac{1}{8c_\rho}(\frac{r}{2})^2,$$

dann liefert (e): $\mathbb{P}(A^c) \leq e^{-t}$. Dies wird zum Beispiel erfüllt durch

$$\frac{1}{24c_\rho}(\frac{r}{2})^2 \geq 2\phi_\rho(r^2), \qquad \frac{1}{24c_\rho}(\frac{r}{2})^2 \geq \sqrt{\frac{2t}{n}} \cdot r, \qquad \frac{1}{24c_\rho}(\frac{r}{2})^2 \geq \frac{16(\rho+1)}{3} \cdot \frac{t}{n},$$

oder äquivalent

$$\frac{r^2}{192c_\rho} \geq \phi_\rho(r^2), \qquad r \geq 96c_\rho\sqrt{\frac{2t}{n}}, \qquad r \geq 16\sqrt{c_\rho(\rho+1)} \cdot \sqrt{\frac{2t}{n}}$$

Die erste Gleichung ist erfüllt, wenn (Hinweis): $r \geq 4 \cdot 192c_\rho\gamma(n)^{1/2}$.

Zusätzlich muss gelten: $\tilde{R}(\delta_0) - \tilde{R}(\delta^*) \leq \frac{1}{8c_\rho}(\frac{r}{2})^2$, d.h.

$$r \geq 2\big(8c_\rho\{\tilde{R}(\delta_0) - \tilde{R}(\delta^*)\}\big)^{1/2}.$$

Zusammenfassung der Ungleichungen liefert die Aussage.

(g) Einsetzen von r aus (f) in die Ungleichung aus (d) liefert (c bezeichnet eine Zahl, die hier nicht genauer bestimmt wird).

$$\begin{aligned}
\tilde{R}(\hat{\delta}) - \tilde{R}(\delta^*) &\leq \{\tilde{R}(\delta_0) - \tilde{R}(\delta^*)\} + \frac{1}{8c_\rho}(\frac{r}{2})^2 \\
&\leq 2\{\tilde{R}(\delta_0) - \tilde{R}(\delta^*)\} + \frac{1}{8c_\rho} \cdot \Big((4 \cdot 192)^2 c_\rho^2\gamma(n) + (96)^2 c_\rho^2 \cdot \frac{2t}{n} + 16^2 c_\rho(\rho+1) \cdot \frac{2t}{n}\Big) \\
&\leq 2\{\tilde{R}(\delta_0) - \tilde{R}(\delta^*)\} + c \cdot \big\{c_\rho\gamma(n) + (c_\rho + \rho + 1) \cdot \frac{t}{n}\big\}.
\end{aligned}$$

Wahl von δ_0 liefert

$$\tilde{R}(\delta_0) - \tilde{R}(\delta^*) = \inf_{\delta \in B(\rho)}\{\tilde{R}(\delta) - \tilde{R}(\delta^*)\}.$$

(h) Zu ändern ist das Vorgehen in (c). Definiere für $\varepsilon > 0$:

$$A := \{Z_r \leq \frac{\varepsilon}{1+\varepsilon}\frac{1}{4c_\rho}(\frac{r}{2})^2\}$$

und fordere $\tilde{R}(\delta_0) - \tilde{R}(\delta^*) \leq \frac{1}{(1+\varepsilon)4c_\rho}(\frac{r}{2})^2$. Dann gilt weiterhin in (d): $D(\tilde{\delta}, \delta_0) \leq \frac{r}{2}$ und somit $D(\hat{\delta}, \delta_0) \leq r$. In (f) folgt:

$$r \geq \max\Big\{..., 2\big((1+\varepsilon)4c_\rho\{\tilde{R}(\delta_0) - \tilde{R}(\delta^*)\}\big)^{1/2}\Big\}$$

und damit in (g):

$$\begin{aligned}
\tilde{R}(\hat{\delta}) - \tilde{R}(\delta^*) &\leq \{\tilde{R}(\delta_0) - \tilde{R}(\delta^*)\} + \frac{\varepsilon}{(1+\varepsilon)4c_\rho}(\frac{r}{2})^2 \\
&\leq \{\tilde{R}(\delta_0) - \tilde{R}(\delta^*)\} + \varepsilon\{\tilde{R}(\delta_0) - \tilde{R}(\delta^*)\} + ... \\
&= (1+\varepsilon) \cdot \{\tilde{R}(\delta_0) - \tilde{R}(\delta^*)\} + ...
\end{aligned}$$

Natürlich sind die Restterme in ... jetzt aber größer, da dort zum Beispiel $\frac{1}{\varepsilon}$ als Faktor vorkommt.

(i) Für jedes $m \in \mathbb{N}$ haben wir in (a)-(g) gesehen:

$$\mathbb{P}\Big(\underbrace{\tilde{R}(\hat{\delta}) - \tilde{R}(\delta^*) \geq 2\{\tilde{R}(\delta_m) - \tilde{R}(\delta^*)\} + c \cdot \{c_\rho \gamma(n) + (c_\rho + \rho + 1) \cdot \frac{t}{n}}_{=:B_m}\Big) \leq e^{-t}.$$

Wegen $\tilde{R}(\delta_m) \downarrow \inf_{\delta \in B(\rho)} \tilde{R}(\delta)$ gilt $B_m \uparrow$ und

$$\bigcup_{m \in \mathbb{N}} B_m = B := \Big\{ \inf_{\delta \in B(\rho)} \tilde{R}(\delta) - \tilde{R}(\delta^*) \geq 2\{\tilde{R}(\delta_m) - \tilde{R}(\delta^*)\} + c \cdot \{c_\rho \gamma(n) + (c_\rho + \rho + 1) \cdot \frac{t}{n}\Big\}.$$

Es folgt

$$\mathbb{P}(B) = \lim_{m \to \infty} \mathbb{P}(B_m) \leq e^{-t}.$$

Aufgabe 2 (Einordnung: Bernstein-Ungleichung und Talagrand-Ungleichung + Separabilität)
Lösung:

(a) Es gilt

$$\sqrt{2tv} \overset{\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}}{\leq} \sqrt{2tn}\sigma + 2\sqrt{tM\mathbb{E}Z} \overset{2\sqrt{ab} \leq \alpha a + \frac{b}{\alpha}}{\leq} \sqrt{2tn}\sigma + \alpha \mathbb{E}Z + \frac{1}{\alpha} tM.$$

Talagrand $\Rightarrow$ Mit WS $\geq 1 - e^{-t}$,

$$Z \leq \mathbb{E}Z + \sqrt{2tv} + \frac{tM}{3} \leq (1+\alpha)\mathbb{E}Z + \sqrt{2tn}\sigma + (\frac{1}{3} + \frac{1}{\alpha})Mt.$$

(b) Nur der Term $\mathbb{E}Z$ wird zusätzlich benötigt. Die anderen beiden Terme $\sqrt{2tn}\sigma$ und tM kommen bei beiden Ungleichungen vor.
Für festes f variiert $\sum_{i=1}^n f(X_i)$ um 0, die Variation wird dann nur mit $\sqrt{2tn}\sigma$ und tM erklärt.
$Z = \sup_{f \in \mathcal{F}} \sum_{i=1}^n f(X_i)$ variiert um $\mathbb{E}Z$ (natürlich gilt i.A. $\mathbb{E}Z > 0$); abgesehen von diesem veränderten Erwartungswert verhält sich die Variation aber wie für festes f. Die 'Komplexität' der Funktionenklasse $\mathcal{F}$ (nicht die groben oberen Schranken σ^2, M) geht daher vor allem in $\mathbb{E}Z$ ein und sorgt nicht für zusätzliche Variation von Z, sondern für eine *Verschiebung* vom Nullpunkt weg. Die eigentliche Schwierigkeit ist oft, eine gute obere Schranke für $\mathbb{E}Z$ zu finden.

(c) $\mathcal{F}_{sep} \subset \mathcal{F} \Rightarrow \sup_{f \in \mathcal{F}} \sum_{i=1}^n f(X_i) \geq \sup_{g \in \mathcal{F}_{sep}} \sum_{i=1}^n g(X_i)$.
Zeige nun $\leq$: Sei $\varepsilon > 0$. Sei $f \in \mathcal{F}$ beliebig. Sei $g \in \mathcal{F}_{sep}$ mit $\|f - g\| \leq \varepsilon$. Dann gilt

$$\sum_{i=1}^n f(X_i) \leq \sum_{i=1}^n g(X_i) + \sum_{i=1}^n \underbrace{|f(X_i) - g(X_i)|}_{\leq C\|f-g\| \leq C\varepsilon} \leq \sum_{i=1}^n g(X_i) + nC\varepsilon.$$

$\Rightarrow$

$$\sup_{f \in \mathcal{F}} \sum_{i=1}^n f(X_i) \leq \sup_{g \in \mathcal{F}_{sep}} \sum_{i=1}^n g(X_i) + nC\varepsilon.$$

Mit $\varepsilon \to 0$ folgt die Behauptung (Achtung: n, C sind während dieser Argumentation fest!)

(d) Wir zeigen, dass für $\delta, \delta' \in B(\rho)$ gilt: $|f_\delta(x,y) - f_{\delta'}(x,y)| \le \|\delta - \delta'\|_K$.
Beachte: $\|\delta\|_\infty \le \|\delta\|_K \le \rho$ (vgl. Lemma 4.24(ii)). Es gilt $f_\delta = \frac{A(\delta)}{B(\delta)}$ mit

$$A(\delta) := (\mathbb{E}\tilde{L}(Y,\delta(X)) - \mathbb{E}\tilde{L}(Y,\delta_0(X))) - (\tilde{L}(y,\delta(x)) - \tilde{L}(y,\delta_0(x))), \qquad B(\delta) := n(D(\delta,\delta_0)^2 + r^2).$$

Daher

$$|f_\delta(x,y) - f_{\delta'}(x,y)| \le \frac{1}{B(\delta)} \cdot \left|A(\delta) - A(\delta')\right| + \frac{|A(\delta')|}{B(\delta)\cdot B(\delta')} \cdot \left|B(\delta) - B(\delta')\right|.$$

Es ist $\frac{1}{B(\delta)} \le \frac{1}{nr^2}$, $|A(\delta)| \le 4(\rho+1)$. Außerdem

$$|A(\delta) - A(\delta')| \le \left|\mathbb{E}\tilde{L}(Y,\delta(X)) - \mathbb{E}\tilde{L}(Y,\delta'(X))\right| + \left|\tilde{L}(y,\delta(x)) - \tilde{L}(y,\delta'(x))\right| \le 2\|\delta - \delta'\|_\infty,$$

und wegen $|D(\delta,\delta_0) - D(\delta',\delta_0)| \le \|\delta - \delta'\|_\infty$, $D(\delta,\delta_0) \le 2\rho$,

$$|B(\delta) - B(\delta')| \le \left|D(\delta',\delta_0)^2 - D(\delta,\delta_0)^2\right| \le \|\delta - \delta'\|_\infty \cdot 4\rho.$$

Zusammengefasst:

$$|f_\delta(x,y) - f_{\delta'}(x,y)| \le \underbrace{\left(\frac{2}{nr^2} + \frac{4(\rho+1)\cdot 4\rho}{(nr^2)^2}\right)}_{=:C} \cdot \underbrace{\|\delta - \delta'\|_\infty}_{\le \|\delta-\delta'\|_K}.$$

Aufgabe 3 (Erwartetes excess Bayes risk anstatt Abweichungsungleichung)
Lösung:

(a) Es gilt

$$\begin{aligned}
\mathbb{E}Z &= \int_0^\infty \mathbb{P}(Z \ge x)dx = \int_0^A \underbrace{\mathbb{P}(Z \ge x)}_{\le 1} dx + \int_A^\infty \mathbb{P}(Z \ge x)dx \\
&\overset{\text{Subst. } x = y + A}{\le} A + \int_0^\infty \mathbb{P}(Z \ge A + y)dy \\
&\overset{\text{Subst. } y = Bt}{=} A + B \cdot \int_0^\infty \underbrace{\mathbb{P}(Z \ge A + Bt)}_{\le g(t)} dt \le A + B \cdot \int_0^\infty g(t)dt.
\end{aligned}$$

(b) Wähle $g(t) = e^{-t}$ (damit $\int_0^\infty g(t)dt = 1$),

$$\begin{aligned}
Z &= \tilde{R}(\hat{\delta}) - \tilde{R}(\delta^*), \\
A &= 2 \inf_{\delta \in B(\rho)} \{\tilde{R}(\delta) - \tilde{R}(\delta^*)\} + c \cdot c_\rho \gamma(n), \\
B &= \frac{c}{n} \cdot (c_\rho + \rho + 1)
\end{aligned}$$

Die Aussage folgt nun aus (a) und Aufgabe 2(g).
(Eine ähnliche Beweistechnik mit Bestrafungsterm ist komplizierter, da im Beweis auch der Bestrafungsterm von t abhängt).

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