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

7. Übungsblatt

Aufgabe 1 (Alternativer Beweis des vereinfachten SVM-Schätzers).
Sei $\mathcal{H}_K$ der RKHS zu einem Mercer-Kern K. In dieser Aufgabe betrachten wir für festes $\rho > 0$ den Schätzer

$$\hat{\delta} \in \mathrm{argmin}_{\delta \in B(\rho)} \tilde{R}_n(\delta), \qquad B(\rho) := \{\delta \in \mathcal{H}_K : \|\delta\|_K \leq \rho\},$$

wobei $\tilde{R}_n(\delta) = \frac{1}{n}\sum_{i=1}^n \tilde{L}(Y_i, \delta(X_i))$ und $\tilde{L}(y,s) = (1-ys)_+$ der Hinge-Verlust ist. Sei $D(\delta, \delta_0) := \mathbb{E}[(\delta(X) - \delta^*(X))^2]^{1/2}$. Sei $\delta_0 \in \mathrm{argmin}_{\delta \in B(\rho)} \tilde{R}(\delta)$, wobei $\tilde{R}(\delta) := \mathbb{E}\tilde{L}(Y, \delta(X))$.

(a) Sei $r > 0$ und $\tilde{\delta} := T\hat{\delta} + (1-T)\delta_0$ mit $T := \frac{r}{r + D(\hat{\delta}, \delta_0)}$. Zeigen Sie, dass

$$\tilde{\delta} \in B(\rho), \qquad D(\tilde{\delta}, \delta_0) \leq r.$$

(b) Folgern Sie, dass

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

wobei $Z_r := \sup_{\delta \in B(\rho), D(\delta,\delta_0) \leq r}\{\tilde{R}(\delta) - \tilde{R}_n(\delta) - (\tilde{R}(\delta_0) - \tilde{R}_n(\delta_0))\}$.
Hinweis: Nutzen Sie, dass $\tilde{L}(y,s)$ konvex in s ist.

(c) Definiere $A := \{Z_r \leq \frac{1}{8c_\rho}(\frac{r}{2})^2\}$. Zeigen Sie, dass auf A und unter der Annahme $\tilde{R}(\delta_0) - \tilde{R}(\delta^*) \leq \frac{1}{8c_\rho}(\frac{r}{2})^2$ gilt:

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

Hinweis: Verwenden Sie die quadratic margin condition $D(\delta, \delta^*)^2 \leq c_\rho\{\tilde{R}(\delta) - \tilde{R}(\delta^*)\}$.

(d) Folgern Sie, dass unter den Bedingungen aus (c) gilt:

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

(e) Nach Lemma 4.27 ist die Abschätzung $\mathbb{E}|Z_r| \leq \phi_\rho(r^2)$ bereits bekannt. Nutzen Sie die Talagrand-Ungleichung Satz 4.29 (und die Folgerung mit $\alpha = 1$), um zu zeigen: Für jedes $t > 0$ gilt

$$\mathbb{P}\Big(Z_r \geq 2\phi_\rho(r^2) + \sqrt{\frac{2t}{n}} \cdot r + \frac{16(\rho+1)}{3} \cdot \frac{t}{n}\Big) \leq e^{-t}$$

Hinweis: Für $\delta \in B(\rho)$ gilt $\|\delta\|_\infty \leq \rho$.

(f) Zeigen Sie, dass mit

$$r = \max\Big\{4 \cdot 192 c_\rho \gamma(n)^{1/2}, \quad 96 c_\rho\sqrt{\frac{2t}{n}}, \quad 16\sqrt{c_\rho(\rho+1)} \cdot \sqrt{\frac{2t}{n}}, \quad 2\big(8c_\rho\{\tilde{R}(\delta_0) - \tilde{R}(\delta^*)\}\big)^{1/2}\Big\}$$

gilt: $\mathbb{P}(A^c) \leq e^{-t}$ und $\tilde{R}(\delta_0) - \tilde{R}(\delta^*) \leq \frac{1}{2c_\rho}(\frac{r}{2})^2$.
Hinweise:

- *Schätzen Sie alle 3 Terme in der Wahrscheinlichkeit in (e) einzeln durch* $\frac{1}{24c_\rho}(\frac{r}{2})^2$ *ab.*
- *Die Gleichung* $\frac{r^2}{192c_\rho} \geq \phi_\rho(r^2)$ *ist erfüllt, wenn* $r^2 \geq 16(192c_\rho)^2\gamma(n)$*, vgl. Lemma 4.31.*

(g) Folgern Sie, dass mit Wahrscheinlichkeit $\geq 1 - e^{-t}$ gilt:

$$\tilde{R}(\hat{\delta}) - \tilde{R}(\delta^*) \leq 2 \inf_{\delta \in B(\rho)} \{\tilde{R}(\delta) - \tilde{R}(\delta^*)\} + c \cdot \big\{c_\rho\gamma(n) + (c_\rho + \rho + 1) \cdot \frac{t}{n}\big\},$$

wobei c eine (möglicherweise große) Zahl ist.
Hinweis: Für $a, b \geq 0$ *gilt* $\max\{a, b\} \leq a + b$.

(h) Diskussion: Wie und wo muss der Beweis modifiziert werden, damit stattdessen mit WS $\geq 1 - e^{-t}$ gilt:

$$\tilde{R}(\hat{\delta}) - \tilde{R}(\delta^*) \leq (1 + \varepsilon)\{\tilde{R}(\delta_0) - \tilde{R}(\delta^*)\} + c(\varepsilon) \cdot \big\{c_\rho\gamma(n) + (c_\rho + \rho + 1) \cdot \frac{t}{n}\big\}$$

wobei $\varepsilon > 0$ beliebig klein und $c(\varepsilon) > 0$ eine Konstante nur abhängig von ε erhalten wird?

(i) Ergänzung: Der Beweis geht (im Grunde unberechtigt) davon aus, dass ein $\delta_0 \in B(\rho)$ existiert, so dass $\tilde{R}(\delta_0) = \inf_{\delta \in B(\rho)} \tilde{R}(\delta)$. Im Allgemeinen kann man nur eine Folge $(\delta_m)_{m \in \mathbb{N}} \subset B(\rho)$ finden mit $\tilde{R}(\delta_m) \downarrow \inf_{\delta \in B(\rho)} \tilde{R}(\delta)$. Der Beweis muss also statt mit δ_0 mit jedem δ_m, $m \in \mathbb{N}$ durchgeführt werden. Fassen Sie kurz zusammen, wo der Beweis modifiziert werden muss.

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

In dieser Aufgabe diskutieren wir kurz die Aussagen und Zusammenhänge zwischen Bernstein- und Talagrand-Ungleichung. Seien $X_i \in \mathcal{X} \subset \mathbb{R}^d$, $i = 1, ..., n$ i.i.d. und $\mathcal{F} \subset \{f : \mathcal{X} \to \mathbb{R} \text{ messbar}\}$ abzählbar. Es gelte $\mathbb{E}f(X) = 0$, $\sup_{f \in \mathcal{F}} \mathrm{Var}(f(X)) \leq \sigma^2$ und $\sup_{f \in \mathcal{F}} \|f\|_\infty \leq M$. Sei $v := n\sigma^2 + 2M\mathbb{E}Z$. Dann gilt für $Z := \sup_{f \in \mathcal{F}} \sum_{i=1}^n f(X_i)$ und $t > 0$:

$$\begin{aligned} \mathbb{P}(Z \geq \mathbb{E}Z + \sqrt{2tv} + \frac{tM}{3}) &\leq e^{-t} \quad \text{(Talagrand)}, \\ \mathbb{P}(\sum_{i=1}^n f(X_i) \geq \sqrt{2tn}\sigma + \frac{tM}{3}) &\leq e^{-t} \quad \text{(Bernstein)}. \end{aligned}$$

(a) Zeigen Sie die Aussage nach Satz 4.29 im Detail: Für $\alpha > 0$ folgt aus der Talagrand-Ungleichung:

$$\mathbb{P}(Z \geq (1 + \alpha)\mathbb{E}Z + \sqrt{2tn}\sigma + (\frac{1}{\alpha} + \frac{1}{3})tM) \leq e^{-t}$$

(b) Vergleichen Sie die Aussagen von Talagrand- und Bernstein-Ungleichung. Welcher Term wird zusätzlich benötigt, um die Variation von $\sup_{f \in \mathcal{F}} \sum_{i=1}^n f(X_i)$ im Gegensatz zu der von $\sum_{i=1}^n f(X_i)$ für festes $f \in \mathcal{F}$ zu erklären? Interpretieren Sie dieses Ergebnis.

(c) Angenommen, $\mathcal{F}$ ist nicht abzählbar, aber (aufgefasst als Teilmenge eines normierten Raumes $(\mathcal{E}, \|\cdot\|)$) separabel, d.h. es gibt eine dichte Teilmenge $\mathcal{F}_{sep} \subset \mathcal{F}$, die dicht in $\mathcal{F}$ liegt. Weiter gebe es $C > 0$, so dass für alle $x \in \mathcal{X}$ und $f, g \in \mathcal{F}$ gilt: $|f(x) - g(x)| \leq C\|f - g\|$. Zeigen Sie, dass dann gilt:

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

(d) Anwendung beim SVM-Beweis (Lemma 4.30): Als Hilbertraum mit abzählbarer Orthonormalbasis ist $\mathcal{H}_K$ separabel. Zeigen Sie mit (c), dass die Talagrand-Ungleichung auf die Menge $\mathcal{F} \subset \{f_\delta(x,y) = \frac{(\mathbb{E}\tilde{L}(Y,\delta(X)) - \mathbb{E}\tilde{L}(Y,\delta_0(X))) - (\tilde{L}(y,\delta(x)) - \tilde{L}(y,\delta_0(x)))}{n(D(\delta,\delta_0)^2 + r^2)} : \delta \in B(\rho)\}$ anwendbar ist.

Aufgabe 3 (Erwartetes excess Bayes risk anstatt Abweichungsungleichung).
In dieser Aufgabe leiten wir aus Abweichungsungleichungen obere Schranken für Erwartungswerte her. Die Herleitung basiert im Wesentlichen auf der Gleichung

$$\mathbb{E}Z = \int_0^\infty \mathbb{P}(Z \geq x)dx$$

für Zufallsvariablen $Z \geq 0$.

(a) Es gelte $\mathbb{P}(Z \geq A + B \cdot t) \leq g(t)$, wobei $\int_0^\infty g(t)dt < \infty$. Zeigen Sie, dass

$$\mathbb{E}Z \leq A + B \cdot \int g(t)dt.$$

(b) Zeigen Sie, dass für den Schätzer aus Aufgabe 2 gilt:

$$\mathbb{E}\tilde{R}(\hat{\delta}) - \tilde{R}(\delta^*) \leq 2 \inf_{\delta \in B(\rho)} \{\tilde{R}(\delta) - \tilde{R}(\delta^*)\} + c \cdot \{c_\rho \gamma(n) + (c_\rho + \rho + 1) \cdot \frac{1}{n}\}.$$

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