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

6. Übungsblatt

Aufgabe 1 (SVM: Zu Kernen gehörige nichtlineare Transformationen).
Es sei $K_p : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}, K_p(x, x') := (1 + x^T x')^p$ der Polynomkern zum Grad $p \in \mathbb{N}$.

(a) Sei zunächst $d = 2$, $p = 2$. Finden Sie eine Abbildung $h : \mathbb{R}^2 \to \mathbb{R}^6$, so dass $K_p(x, x') = h(x)^T h(x')$.

(b) Seien nun p, d beliebig. Finden Sie eine Abbildung $h : \mathbb{R}^d \to \mathbb{R}^m$, so dass $K_p(x, x') = h(x)^T h(x')$. Geben Sie eine obere Schranke für $m \in \mathbb{N}$ an.
Hinweis: Definieren Sie $\tilde{x} := (1, x)$, $\tilde{x}' := (1, x')$ und schreiben Sie $K_p(x, x') = (\tilde{x}, \tilde{x}')^p$.

Sei nun $K_\gamma : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}, K_\gamma(x, x') = \exp(-\gamma \|x - x'\|_2^2)$ der Gaußkern zum Abstieg γ.

(c) Sei $d = 1$. Finden Sie eine Abbildung $h : \mathbb{R} \to \ell^2$, so dass $K_\gamma(x, x') = h(x)^T h(x')$.

(d) Sei d beliebig. Wie sieht h jetzt aus?

Aufgabe 2 (Analyse der Raten in Satz 4.22).
Sei $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ ein Mercer-Kern. In der Orakel-Ungleichung von Satz 4.22 ist die Rate des Schätzfehlers im Wesentlichen durch

$$\gamma(n) = \frac{1}{\sqrt{n}} \inf_{a \in \mathbb{N}} \{ \frac{a}{\sqrt{n}} + \sqrt{\sum_{j > a} \gamma_j} \}$$

gegeben (hier der Einfachheit halber ohne η_1), wobei γ_j die Eigenwerte des Integraloperators $T_K : L^2(\mathbb{P}^X) \to L^2(\mathbb{P}^X), (T_K g)(x) := \int K(x, x') g(x') \, \mathrm{d}\mathbb{P}^X(x')$ sind. In dieser Aufgabe leiten wir (schlechte) obere Schranken für verschiedene Strukturen von γ_j her.

(a) Es gebe $C > 0$ mit $\gamma_j = 0$ für $j > C$. Zeigen Sie, dass $\gamma(n) \leq \frac{C}{n}$.

(b) Es gebe $C > 0$ mit $\gamma_j \leq C j^{-\alpha}$ ($\alpha > 1$). Zeigen Sie, dass $\gamma(n) \leq \tilde{c}_\alpha \cdot (\frac{1}{n} + C^{\frac{1}{\alpha+1}} n^{-\frac{\alpha}{\alpha+1}})$, wobei $\tilde{c}_\alpha$ eine Konstante ist, die nur von α abhängt.
Hinweis: Nutzen Sie ohne Beweis, dass $\sum_{j>a} j^{-\alpha} \leq c_\alpha a^{-\alpha+1}$ mit einem $c_\alpha > 0$.

(c) Es gebe $C > 0$ mit $\gamma_j \leq C\rho^j$ ($\rho \in (0, 1)$). Zeigen Sie, dass $\gamma(n) \leq c_\rho \cdot (\frac{1}{n} + 2\frac{\log(nC)}{nC^{1/2}})$, wobei c_ρ eine Konstante ist, die nur von ρ abhängt.

Falls $\mathcal{X} = [0, 1]^d$ und $\mathbb{P}^X = U[0, 1]^d$ die Gleichverteilung auf $[0, 1]^d$, so ist $(T_K g)(x) = \int K(x, x') g(x') dx'$. Sei $K(x, x') = h(x)^T h(x')$ mit geeignetem $h : \mathbb{R}^d \to \ell^2$.

(d) Zeigen Sie: Ist $(h_k)_{k \in \mathbb{N}}$ orthogonal in $L^2(\lambda)$, so ist $\gamma_k = \int h_k(x)^2 dx$.

(e) Argumentieren Sie, dass für den (normierten) Polynomkern $K_p(x, x')$ gilt: $\gamma(n) \leq (d+1)^p$. *Nutzen Sie (a),(d) und A1(b).*

Aufgabe 3 (Margin condition der SVM unter der allgemeinen Rauschbedingung).

Zum Beweis der quadratic margin condition der SVM in Lemma 4.24 wurde angenommen, dass $|\eta(x) - \frac{1}{2}| \geq \eta_0$ mit einem $\eta_0 > 0$, d.h. die Rauschbedingung muss mit $q = \infty$ erfüllt sein. Angenommen, die Rauschbedingung gilt nur mit $q \in (1, \infty)$, d.h. es gibt $C > 0$ mit

$$\forall t > 0: \quad \mathbb{P}(|\eta(X) - \frac{1}{2}| \leq t) \leq Ct^q.$$

Zeigen Sie, dass dann für $\delta : \mathcal{X} \to \mathbb{R}$ mit $\|\delta\|_\infty \leq \rho$ und $\delta^*(x)$ die Bayes-Regel zu $\tilde{L}(y, s) = (1 - ys)_+$ gilt:

$$\mathbb{E}[(\delta(X) - \delta^*(X))^2] \leq \tilde{c}_\rho \cdot \big\{\tilde{R}(\delta) - \tilde{R}(\delta^*)\big\}^{\frac{q}{q+1}},$$

wobei $\tilde{c}_\rho > 0$ eine Konstante ist, die nur von ρ, η_0, η_1, C abhängt.
Hinweise:

- Es wurde in Lemma 4.24 bereits gezeigt, dass $A(x) = \frac{(\delta(x)-\delta^*(x))^2}{\mathbb{E}[\tilde{L}(Y,\delta(X))-\tilde{L}(Y,\delta^*(X))|X=x]} \leq c_\rho(\eta_0) = 2(\frac{\rho}{\eta_1} + \frac{1}{\eta_0})$.
- Fügen Sie $1 = \mathbb{1}_{\{|\eta(X)-\frac{1}{2}|>t\}} + \mathbb{1}_{\{|\eta(X)-\frac{1}{2}|\leq t\}}$ in $\mathbb{E}[(\delta(X) - \delta^*(X))^2]$ ein, wählen Sie t geeignet (vgl. Beweis der verbesserten Risiko-Übertragungsformel, Satz 3.24).
- Nutzen Sie am Ende, dass $\tilde{R}(\delta) - \tilde{R}(\delta^*) \leq (\rho + 1)^{\frac{1}{q}} (\tilde{R}(\delta) - \tilde{R}(\delta^*))^{\frac{q}{q+1}}$.

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