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

6. Übungsblatt

Aufgabe 1 (SVM: Zu Kernen gehörige nichtlineare Transformationen)
Lösung:

(a) Es gilt

$$\begin{aligned} K_2(x,x') &= (1+x^Tx')^2 \\ &= (1+x_1x_1'+x_2x_2')^2 = 1 + x_1^2(x_1')^2 + x_2^2(x_2')^2 + 2x_1x_1' + 2x_2x_2' + 2x_1x_2x_1x_2' \\ &= h(x)^Th(x') \end{aligned}$$

mit

$$h(x) = (1, x_1^2, x_2^2, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1x_2).$$

(b) Mit $\tilde{x} := (1,x)$, $\tilde{x}' := (1,x')$ gilt

$$\begin{aligned} K_p(x,x') &= (\tilde{x},\tilde{x}')^p = \Big(\sum_{k=1}^{d+1} \tilde{x}_k\tilde{x}_k'\Big)^p \\ &= \sum_{i_1,...,i_p=1}^{d+1} (\tilde{x}_{i_1}\cdot ... \cdot x_{i_p})\cdot(\tilde{x}_{i_1}'\cdot ... \cdot x_{i_p}') \\ &= h(x)^Th(x'), \end{aligned}$$

wobei $h(x) = (\tilde{x}_{i_1}\cdot ...\tilde{x}_{i_p})_{i_1,...,i_p=1,...,d+1}$.
Das ist in keiner Hinsicht eine optimale Darstellung, h enthält sogar viele Komponenten doppelt. Allerdings genügt es, um einen Eindruck von den zu K gehörenden nichtlinearen Transformationen zu bekommen.
Offensichtlich gilt hier $m \le |\{(i_1,...,i_p) : i_1,...,i_p \in \{1,...,d+1\}\}| = (d+1)^p$.

(c) Es gilt

$$\begin{aligned} K_\gamma(x,x') &= \exp(-\gamma(x-x')^2) = \exp(-\gamma x^2 + 2\gamma xx' - \gamma(x')^2) \\ &= e^{-\gamma(x^2+(x')^2)}\sum_{k=0}^{\infty}\frac{(2\gamma xx')^k}{k!} \\ &= \sum_{k=0}^{\infty}\Big(\frac{(2\gamma)^{k/2}}{(k!)^{1/2}}x^ke^{-\gamma x^2}\Big)\cdot\Big(\frac{(2\gamma)^{k/2}}{(k!)^{1/2}}(x')^ke^{-\gamma(x')^2}\Big), \end{aligned}$$

d.h. wähle $h(x) = \big(\frac{(2\gamma)^{k/2}}{(k!)^{1/2}}x^ke^{-\gamma x^2}\big)_{k\in\mathbb{N}_0}$.
Die nichtlinearen Transformationen entsprechen also Polynomen x^k ($k \in \mathbb{N}_0$), deren Oszillation mit dem Faktor $e^{-\gamma x^2}$ abgeschwächt wird.

(d) Sei H die Funktion h für den eindimensionalen Gaußkern aus (c). Dann gilt

$$\begin{aligned} K_\gamma(x,x') &= e^{-\gamma\|x-x'\|_2^2} = \prod_{j=1}^d e^{-\gamma(x_j-x_j')^2} = \prod_{j=1}^d \sum_{k=0}^\infty H_k(x_k)H_k(x_k') \\ &= \sum_{k_1,\dots,k_d=0}^\infty (H_{k_1}(x_{k_1})\cdot\ldots\cdot H_{k_d}(x_{k_d}))\cdot(H_{k_1}(x'_{k_1})\cdot\ldots\cdot H_{k_d}(x'_{k_d})), \end{aligned}$$

d.h. wähle

$$h(x) = (H_{k_1}(x_{k_1})\cdot\ldots\cdot H_{k_d}(x_{k_d}))_{k_1,\dots,k_d\in\mathbb{N}_0} = \Big(\frac{(2\gamma)^{(k_1+\ldots+k_d)/2}}{(k!)^{d/2}} x_1^{k_1}\cdot\ldots\cdot x_d^{k_d} e^{-\gamma\|x\|_2^2}\Big)_{k_1,\dots,k_d\in\mathbb{N}_0}.$$

Aufgabe 2 (Analyse der Raten in Satz 4.22)
Lösung:

(a) Wähle $a = C$ in der Definition von $\gamma(n)$.
Voraussetzung $\Rightarrow \sum_{k>a}\gamma_j = 0 \Rightarrow \gamma(n) \le \frac{C}{n}$.

(b) Hinweis $\Rightarrow$ Es gilt für alle $a\in\mathbb{N}$:

$$\gamma(n) \le \frac{1}{\sqrt{n}}\Big\{\frac{a}{\sqrt{n}} + C^{1/2}c_\alpha^{1/2}a^{(1-\alpha)/2}\Big\}.$$

Ableiten der rechten Seite nach a: $\frac{1}{\sqrt{n}} - C^{1/2}c_\alpha^{1/2}\frac{\alpha-1}{2}a^{-\frac{\alpha+1}{2}}$. Dies legt Wahl

$$a = \lceil (nC)^{\frac{1}{\alpha+1}}\rceil$$

nahe (Konstanten c_α, α haben wir bei der Wahl nicht berücksichtigt, da sie nur Einfluss auf die Konstante $\tilde{c}_\alpha$ nehmen).
$\Rightarrow$

$$\gamma(n) \le \frac{1}{\sqrt{n}}\Big\{\frac{1}{\sqrt{n}} + n^{\frac{1-\alpha}{2(\alpha+1)}}C^{\frac{1}{\alpha+1}} + c_\alpha^{1/2}n^{\frac{1-\alpha}{2(\alpha+1)}}C^{\frac{1}{\alpha+1}}\Big\} \le \tilde{c}_\alpha\cdot\Big(\frac{1}{n} + C^{\frac{1}{\alpha+1}}n^{-\frac{\alpha}{\alpha+1}}\Big).$$

(c) Prinzip wie bei (b): Es gilt für alle $a\in\mathbb{N}$:

$$\gamma(n) \le \frac{1}{\sqrt{n}}\Big\{\frac{a}{\sqrt{n}} + C^{1/2}\frac{1}{(1-\rho)^{1/2}}\rho^{a/2}\Big\}.$$

Wähle $a = \lceil 2\log_\rho((nC)^{-1})\rceil$. $\Rightarrow$

$$\gamma(n) \le \frac{1}{\sqrt{n}}\Big\{\frac{1}{\sqrt{n}} + 2\log_\rho((nC)^{-1}) + C^{1/2}\frac{1}{(1-\rho)^{1/2}}(nC)^{-1}\Big\} \le c_\rho\cdot\Big(\frac{1}{n} + 2\frac{\log(nC)}{nC^{1/2}}\Big).$$

(d) Für h_k gilt

$$(T_K h_k)(x) = \int K(x,x')h_k(x')dx' = \sum_{l=1}^\infty h_l(x)\int h_l(x')h_k(x')dx' \overset{\text{orthog.}}{=} h_k(x)\underbrace{\int h_k(x')^2dx'}_{=:\gamma_k},$$

d.h. h_k ist Eigenfunktion von T_K zum Eigenwert $\gamma_k = \int h_k(x')^2dx'$.

(e) Aus (d) folgt: T_K hat nicht mehr Eigenwerte $\neq 0$ als die Funktion h Komponenten hat.
A1(b) $\Rightarrow$ h hat höchstens $(d+1)^p$ Komponenten.
(a) mit $C = (d+1)^p \Rightarrow \gamma(n) \leq \frac{(d+1)^p}{n}$.

Aufgabe 3 (Margin condition der SVM unter der allgemeinen Rauschbedingung)
Lösung: Es gilt für $t \in (0, \frac{1}{2})$:

$$\begin{aligned}\mathbb{E}[(\delta(X)-\delta^*(X))^2] &= \mathbb{E}[\underbrace{(\delta(X)-\delta^*(X))^2}_{\leq A(x)\cdot\{\mathbb{E}[\tilde{L}(Y,\delta(X))-\tilde{L}(Y,\delta^*(X))|X=x]\}} \underbrace{\mathbb{1}_{\{|\eta(X)-\frac{1}{2}|>t\}}}_{\leq 1}] \\ &\qquad + \mathbb{E}[\underbrace{(\delta(X)-\delta^*(X))^2}_{\leq \|\delta-\delta^*\|_\infty^2 \leq (\rho+1)^2} \mathbb{1}_{\{|\eta(X)-\frac{1}{2}|\leq t\}}] \\ &\leq c_\rho(t)\{\tilde{R}(\delta)-\tilde{R}(\delta^*)\} + (\rho+1)^2 \cdot \mathbb{P}(|\eta(X)-\frac{1}{2}| \leq t) \\ &\leq \frac{2\rho}{\eta_1}\{\tilde{R}(\delta)-\tilde{R}(\delta^*)\} + \frac{2}{t}\{\tilde{R}(\delta)-\tilde{R}(\delta^*)\} + C(\rho+1)^2 t^q.\end{aligned}$$

Wähle $t = \{\tilde{R}(\delta)-\tilde{R}(\delta^*)\}^{\frac{1}{q+1}}$. Damit:

$$\begin{aligned}\mathbb{E}[(\delta(X)-\delta^*(X))^2] &\leq \frac{2\rho}{\eta_1}\{\tilde{R}(\delta)-\tilde{R}(\delta^*)\} + \big(2 + C(\rho+1)^2\big)\{\tilde{R}(\delta)-\tilde{R}(\delta^*)\}^{\frac{q}{q+1}} \\ &\overset{\text{Hinweis 3}}{\leq} 2\{\rho \underbrace{(\rho+1)^{1/q}}_{\leq \rho+1} + 2 + C(\rho+1)^2\}\{\tilde{R}(\delta)-\tilde{R}(\delta^*)\}^{\frac{q}{q+1}}.\end{aligned}$$

Es folgt die Behauptung.

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