Computational Learning Theory

Sommersemester 2017

Mario Holldack, M. Sc.
Prof. Dr. Georg Schnitger
Hannes Seiwert, M. Sc.

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Institut für Informatik
AG Theoretische Informatik

Übungsblatt 10

Ausgabe: 26.06.2017
Abgabe: 03.07.2017

Aufgabe 10.1 *Polynomielle Kerne* (5 Punkte)

Sei $d \geq 2$. Geben Sie eine Feature-Funktion ϕ an, sodass $\langle x, z\rangle^d = \langle \phi(x), \phi(z)\rangle$ für $x, z \in \mathbb{R}^n$ gilt.

Aufgabe 10.2 *Noch mehr Kerne* (4 + 2 + 2 + 6 Punkte)

Zeigen Sie:

a) Für natürliche Zahlen x, z ist $K(x,z) := \min(x,z)$ ein Kern.

Hinweis: Geben Sie eine geeignete (möglicherweise unendlich-dimensionale) Feature-Funktion an.

b) Sei $K(x,z) = \langle \phi(x), \phi(z)\rangle$ ein Kern für eine Feature-Funktion ϕ. Dann ist auch K' ein Kern, wobei

$$K'(x,z) := \begin{cases} 0 & , \ K(x,x) = 0 \text{ oder } K(z,z) = 0 \\ \frac{K(x,z)}{\sqrt{K(x,x)K(z,z)}} & , \ \text{sonst} \end{cases}$$

c) Sei K ein Kern mit $\mathrm{Bild}(K) \subseteq [0,1)$. Dann ist auch $K'(x,z) := \frac{1}{1-K(x,z)}$ ein Kern.

d) Sei $U \subseteq \mathbb{N}$ ein endliches Universum. Seien $x, z \subseteq U$ zwei nichtleere Mengen, dargestellt durch ihre Inzidenzvektoren. Dann ist $K(x,z) := \frac{|x \cap z|}{|x \cup z|}$ ein Kern.

Hinweis: Zeigen Sie mithilfe des Teils c), dass $\frac{1}{|x \cup z|} = \frac{1}{|U| - \ldots}$ ein Kern ist, und nutzen Sie weitere Abschlusseigenschaften von Kernen.

Kommentar: Dieser Kern ist auch als Jaccard-Koeffizient bekannt und wird häufig als Ähnlichkeitsmaß für Mengen verwendet.

Aufgabe 10.3 *Representer-Theorem* (5 Punkte)

In der Vorlesung wurde in der Bestimmung eines ERM-Vektors w für eine klassifizierte Beispielmenge $S = \{x_1, \ldots, x_s\} \subseteq \mathbb{R}^n$ angenommen, dass w als Linearkombination der Feature-Vektoren $\phi(x_1), \ldots, \phi(x_s)$ für die Feature-Funktion $\phi \colon \mathbb{R}^n \to \mathcal{H}$ dargestellt werden kann.

Wir liefern hier eine Begründung für sehr allgemeine Formen des Losses mit dem sog. Representer-Theorem: Gegeben sei eine Loss-Funktion *Loss* und eine monoton steigende (Regularisierungs)-Funktion R. Das folgende Minimierungsproblem sei zu lösen:

$$\min_{w \in \mathcal{H}} \Big[\mathit{Loss}\big(\langle w, \phi(x_1)\rangle, \ldots, \langle w, \phi(x_s)\rangle \big) + R(\|w\|) \Big] \tag{1}$$

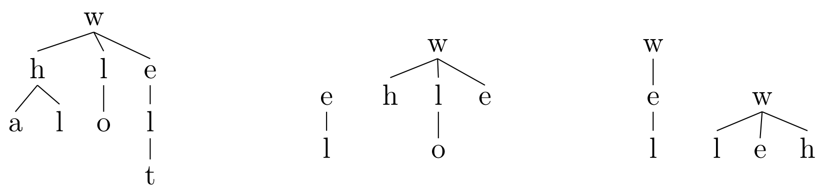
Zeigen Sie: Es gibt eine optimale Lösung w der Form $w = \sum_{i=1}^{s} \alpha_i \phi(x_i)$ mit Koeffizienten $\alpha_i \in \mathbb{R}$. Mit anderen Worten, es gibt ein optimales w in dem durch die Feature-Vektoren $\phi(x_i)$ aufgespannten Unterraum.

Bitte wenden!

Aufgabe 10.4 *Kerne für Bäume* (8 Punkte)

Wir betrachten gewurzelte, geordnete, markierte Bäume, d.h. jeder Knoten hat eine Markierung und die Kinder jedes Knotens sind von links nach rechts angeordnet. Für einen *Teilbaum* T' eines Baumes T müssen alle Knoten von T' entweder alle Kinder oder gar keine besitzen und natürlich in der Ordnung und Markierung der Knoten übereinstimmen.

Beispiel:



Links: ein gewurzelter, geordneter, markierter Baum T. **Mitte:** zwei Teilbäume von T.
Rechts: zwei Bäume, die *keine* Teilbäume von T sind.

Es sei $\mathcal{X}$ die Menge aller gewurzelter, geordneter, markierter Bäume. Wir definieren einen Kern $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ wie folgt: Sei $S_1, S_2, \ldots$ eine Aufzählung aller möglichen Bäume in $\mathcal{X}$. Für einen Baum T gibt $\phi(T)_i$ an, wie oft S_i als Teilbaum von T auftritt. (Wir arbeiten hier also mit unendlich-dimensionalen Feature-Vektoren.)
Schließlich setzen wir $K(T_1, T_2) = \langle \phi(T_1), \phi(T_2) \rangle = \sum_{i=1}^{\infty} \phi(T_1)_i \, \phi(T_2)_i$.

Zeigen Sie, wie K effizient (in Abhängigkeit von den Knotenzahlen) berechnet werden kann.

Hinweis: Sei v_1 ein Knoten aus T_1 und v_2 ein Knoten aus T_2. Bestimmen Sie rekursiv die Anzahl gemeinsamer Teilbäume in T_1 und T_2, die v_1 bzw. v_2 als Wurzeln besitzen.

Kommentar: Solche Kerne für Bäume finden z. B. in der Computerlinguistik bzw. der maschinellen Sprachverarbeitung Anwendung, um die Ähnlichkeit zweier Syntaxbäume festzustellen.
Ähnliche Kerne für allgemeine Graphen werden z. B. in der Chemie bei der Erkennung von Ähnlichkeiten in Molekülstrukturen angewandt. Allerdings ist bei Graph-Kernen Vorsicht geboten: Zu bestimmen, ob ein bestimmter Teilgraph enthalten ist, ist NP-hart. Ein möglicher Ausweg: Betrachte nur sehr einfache Teilgraphen, z. B. Pfade.