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 4

Ausgabe: 15.05.2017
Abgabe: 22.05.2017

Aufgabe 4.1 *Schnitte von Halbräumen* (6 Punkte)

Zeigen Sie, dass das schwache Konsistenzproblem für die Konzeptklasse aller Schnitte zweier Halbräume NP-hart ist.

Hinweis: Reduzieren Sie SetSplitting auf das schwache Konsistenzproblem. Im SetSplitting-Problem sind ein endliches Universum $U = \{1, \ldots, n\}$ und Teilmengen $T_1, \ldots, T_m \subseteq U$ gegeben. Zu entscheiden ist, ob eine Partition $U = U_1 \dot\cup U_2$ existiert, sodass $T_i \not\subseteq U_j$ für alle $i = 1, \ldots, m$ und $j = 1, 2$ gilt.

Für eine Eingabe $U, T_1, \ldots, T_m$ konstruieren Sie die transformierte Eingabe, d. h. eine klassifizierte Beispielmenge $S \subseteq \mathbb{R}^n$ wie folgt:

- Für jedes $u \in U$ klassifiziere den u-ten kanonischen Einheitsvektor e_u negativ.
- Für jedes T_i klassifiziere den Vektor $v_{T_i} := \frac{1}{|T_i|} \sum_{t \in T_i} e_t$ positiv.

Aufgabe 4.2 *k-Term-DNF und k-KNF* (4 + 4 Punkte)

In der Vorlesung wurde gezeigt, dass k-Term-DNFs mit der Hypothesenklasse k-Term-DNF nicht effizient PAC-lernbar sind (sofern NP $\not\subseteq$ RP). Aushilfe schafft hier die Wahl einer anderen (größeren) Hypothesenklasse.

a) Zeigen Sie: Für jede k-Term-DNF φ existiert eine äquivalente k-KNF ψ.

b) Sei $k \in \mathbb{N}$ fest. Beschreiben Sie einen effizienten Algorithmus, der das starke Konsistenzproblem für die Hypothesenklasse k-KNF löst, und bestimmen Sie die Laufzeit.

Bitte wenden!

Aufgabe 4.3 *Schwierigkeit des ERM-Problems im agnostischen Fall* (5 + 5 Punkte)

Wir wissen bereits, wie man das Konsistenzproblem für Monome und Halbräume effizient lösen kann. Nun betrachten wir das Lernen im agnostischen Fall für den 0-1-Loss: Eine Menge S von klassifizierten Beispielen ist gegeben. Eine Hypothese mit der geringsten Anzahl von Fehlklassifikationen – unter allen Hypothesen in der Hypothesenklasse – ist zu bestimmen.

a) Betrachten Sie die Hypothesenklasse $\textsc{Monoton-Monom}_n$ aller monotonen Monome über n Variablen. (Ein monotones Monom ist eine Konjunktion positiver Literale.)
Zeigen Sie: Das schwache ERM-Problem für $\textsc{Monoton-Monom} = (\textsc{Monoton-Monom}_n)_{n\in\mathbb{N}}$ ist NP-hart.

b) Betrachten Sie die Hypothesenklasse $\textsc{Halbraum}_n$, d.h. die Hypothesen sind $\{x\in\mathbb{R}^n : \langle w|x\rangle \geq \tau\}$ für einen Vektor $w \in \mathbb{R}^n$ und einen Schwellwert $\tau \in \mathbb{R}$.
Zeigen Sie: Das schwache ERM-Problem für $\textsc{Halbraum} = (\textsc{Halbraum}_n)_{n\in\mathbb{N}}$ ist NP-hart.

Tipp: Verwenden Sie nur Einheitsvektoren bzgl. der Summennorm $\|\cdot\|_1$ als Beispiele.

Hinweis: Sie können jeweils $\textsc{VertexCover}$ auf das ERM-Problem reduzieren. Wählen Sie jeweils n als die Anzahl der Knoten $|V|$.

Folgende Vorgehensweise ist zu empfehlen: Stellen Sie Knoten durch positive Beispiele und Kanten durch negative Beispiele dar – oder umgekehrt. Sie können Beispiele mehrfach erstellen und dadurch eine Gewichtung der Beispiele simulieren. Wählen Sie die Gewichte der „Kanten-Beispiele" hinreichend groß, um zu gewährleisten, dass jede Lösung des ERM-Problems Fehlklassifizierungen nur auf „Knoten-Beispielen" machen kann.

Beschreiben Sie jeweils Ihre Transformation und führen Sie den Nachweis der Äquivalenz. Eine Begründung der Laufzeit ist nicht nötig.

Aufgabe 4.4 *Totale Ordnungen* (4 + 4 Punkte)

Gegeben sei eine endliche Menge Ω und der Beispielraum $X = \Omega^2$. Als Konzepte (und Hypothesen) betrachten wir alle totalen Ordnungen über Ω. Wenn $\leq_c$ das Zielkonzept ist, wird ein Beispiel $(i, j) \in X$ genau dann mit 1 klassifiziert, wenn $i \leq_c j$ gilt.

a) Beschreiben Sie einen effizienten Algorithmus, der das starke Konsistenzproblem für totale Ordnungen löst, und analysieren Sie die Laufzeit.

b) Wir betrachten jetzt den agnostischen Fall: Jedes Beispiel kann unabhängig vom Zielkonzept mit 0 oder 1 klassifiziert vorliegen.
Zeigen Sie, dass das schwache ERM-Problem für totale Ordnungen NP-hart ist.

Hinweis: Reduzieren Sie $\textsc{FeedbackArcSet}$ auf das ERM-Problem.

Im $\textsc{FeedbackArcSet}$-Problem ist ein gerichteter Graph G und eine Zahl k gegeben und es ist zu entscheiden, ob die Herausnahme von höchstens k Kanten genügt, um alle Kreise zu zerstören.

$$\textsc{FeedbackArcSet} = \big\{(G=(V,E), k) : \exists E' \subseteq E \text{ sodass } |E'| \leq k \text{ und } (V, E\backslash E') \text{ ist azyklisch}\big\}$$