Computational Learning Theory

Sommersemester 2018

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

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Institut für Informatik
AG Theoretische Informatik

Übungsblatt 7

Ausgabe: 28.05.2018
Abgabe: 04.06.2018

Hinweis: Beachten Sie beim Bearbeiten der Aufgaben 7.1 bis 7.4, dass der Perzeptron-Algorithmus aus der Vorlesung und die dazugehörigen Analysen (Satz 8.18 und Satz 8.20) mit Beispielen $(x, 1) \in \mathrm{Def} \times \{1\}$ und Hypothesen der Form $\mathrm{sign}(\langle w, x\rangle)$ statt $\mathrm{sign}(\langle w, x\rangle + t)$ arbeiten.

Aufgabe 7.1 *Perzeptron terminiert nicht immer* (4 + 4 = 8 Punkte)

In der Vorlesung wurde die Frage aufgeworfen, ob der Perzeptron-Algorithmus immer hält. Konstruieren Sie für die beiden folgenden Zielfunktionen eine unendliche Folge $x^{(0)}, x^{(1)}, x^{(2)}, \ldots$ von Gegenbeispielen.

a) Die Zielfunktion sei $f : \mathbb{R}^2 \to \{-1, 1\}$ mit $f(x) = \mathrm{sign}(x_1)$.

b) Die Zielfunktion sei $g : \{0, 1\}^2 \to \{-1, 1\}$ mit $g(x_1, x_2) = \mathrm{sign}\big((x_1 \oplus x_2) - \frac{1}{2}\big)$.

Aufgabe 7.2 *Gute Lehrer für Perzeptron* ($4 + (4{+}6^*) = 8{+}6^*$ Punkte)

Wir nehmen einen Lehrer an, der den Perzeptron-Algorithmus mit möglichst hilfreichen Gegenbeispielen von der Form $(x, 1)$ bzw. $(x, y, 1)$ unterstützt.

a) Wie „wenige“ Gegenbeispiele reichen einem guten Lehrer im besten Fall aus, damit der Perzeptron-Algorithmus die Threshold-Funktion $f_k : \{0, 1\}^n \to \{-1, 1\}$ mit $1 \le k < \frac{n}{2}$ und

$$f_k(x_1, \ldots, x_n) = 1 \iff x_1 \vee \cdots \vee x_k \text{ ist wahr}$$

lernt?

b) i) Wie viele Gegenbeispiele sind auch für den besten Lehrer notwendig, damit der Perzeptron-Algorithmus die Threshold-Funktion $\mathrm{COMP}_n : \{0, 1\}^n \times \{0, 1\}^n \to \{-1, 1\}$ mit

$$\mathrm{COMP}_n(x, y) = \mathrm{sign}\left(\sum_{i=1}^{n} x_i 2^{i-1} - \sum_{i=1}^{n} y_i 2^{i-1}\right)$$

lernt? Eine asymptotische Abschätzung genügt.

ii*) Wie viele Gegenbeispiele reichen im besten Fall aus? Konstruieren Sie wie in a) eine möglichst kurze Folge von Gegenbeispielen (für allgemeines n oder für $n = 2$).

Hinweis: Die untere Schranke aus der Aufgabe 6.3 c) ii) behandelt einen *bösartigen* Lehrer. Können Sie dennoch die Beweisidee recyceln und auf einen hilfreichen Lehrer anwenden?

Bitte wenden!

Aufgabe 7.3 *Vorverarbeitung bei linearer Nichttrennbarkeit* $(4 + 4 = 8$ Punkte)

Bei linearer Trennbarkeit der positiven und negativen Beispiele kann die Klassifikation mithilfe von Threshold-Funktionen gelernt werden. Sind die Beispiele jedoch nicht linear trennbar, so sind wir mit einem schwierigen Problem konfrontiert (vgl. Aufgabe 4.3 a)). In dieser Aufgabe werden wir dieses Problem umgehen, indem wir die Beispiele erst transformieren und dann den Perzeptron-Algorithmus anwenden.

a) Sei $n \in \mathbb{N}_{>0}$ und $X = \{0,1\}^n$. Sei $p : X \to \{0,1\}$ die Paritätsfunktion mit

$$p(x_1, \dots, x_n) = x_1 \oplus x_2 \oplus \cdots \oplus x_n = \sum_{i=1}^{n} x_i \bmod 2.$$

Für alle $t \in \{1, \dots, n\}$ sei $T_t : X \to \{0,1\}$ durch

$$T_t(x_1, \dots, x_n) = 1 \;:\Longleftrightarrow\; \sum_{i=1}^{n} x_i \geq t$$

definiert. Betrachten Sie nun die Transformation $T : X \to \{0,1\}^n$ mit

$$T(x) = \big(T_1(x), \dots, T_n(x)\big),$$

die jedem Beispiel $x \in X$ ein transformiertes Beispiel $T(x)$ zuordnet. Zeigen Sie, dass die transformierten Beispiele linear trennbar sind.

b) Seien $n, r, \tilde{R} \in \mathbb{N}_{>0}$ mit $r \leq \tilde{R} \leq \frac{n}{2}$ und $X = \{-1, 0, 1\}^n$. Sei außerdem

$$K^n_{r,\tilde{R}} := \left\{ (x_1, \dots, x_n) \in X \;:\; \sum_{i=1}^{n} |x_i| \leq r \leq \tilde{R} \right\}$$

die ganzzahlige Kugel mit dem Radius r und dem Mittelpunkt $0 \in \{-1,0,1\}^n$ in der Manhattan-Metrik. Geben Sie eine nichttriviale[1] Transformation $T : X \to \mathbb{R}$ an, sodass die transformierten Beispiele linear trennbar sind. Wie viele Gegenbeispiele reichen für den Perzeptron-Algorithmus auf der transformierten Beispielmenge aus?

Hinweis: Beachten Sie, dass der Margin der transformierten Beispiele strikt größer als Null sein sollte.

Die Transformation der Beispiele spielt für Support-Vektor-Maschinen und neuronale Netzwerke eine zentrale Rolle, siehe auch `https://colah.github.io/posts/2014-03-NN-Manifolds-Topology`.

Aufgabe 7.4 *Perzeptron für* SetSplitting $(2 + 2 + 4 = 8$ Punkte)

In der Aufgabe 4.1 a) haben wir gezeigt, dass das schwache Konsistenzproblem für den Schnitt von Halbräumen NP-hart ist, indem wir SetSplitting auf das schwache Konsistenzproblem reduziert haben. Für eine Eingabe $(U, T_1, \dots, T_m)$ mit $U = \{1, \dots, n\}$ und $\emptyset \neq T_1, \dots, T_m \subseteq U$ haben wir die Beispielfolge $S \subseteq \mathbb{R}^n$ wie folgt konstruiert:

- Für jedes Element $u \in U$ klassifiziere $x^{(u)} := e_u$ negativ.
- Für jede Teilmenge T_i klassifiziere den Vektor $x^{(T_i)} := \frac{1}{|T_i|} \sum_{t \in T_i} e_t$ positiv.

Gehen Sie im Folgenden davon aus, dass $(U, T_1, \dots, T_m)$ eine Ja-Instanz ist.

a) Begründen Sie kurz, weshalb S nicht linear trennbar ist.

b) Wie viele Beispiele werden von $\operatorname{sign}\left(\sum_{i=1}^{n} x_i - \frac{1}{2}\right)$ bzw. $\operatorname{sign}\left(-\sum_{i=1}^{n} x_i + \frac{1}{2}\right)$ falsch klassifiziert?

c) Wie viele Aktualisierungen führt der Perzeptron-Algorithmus höchstens durch, wenn er nur Beispiele aus S – ohne Zurücklegen – erhält, und zwar unabhängig davon, in welcher Reihenfolge die Beispiele vorgelegt werden?

[1] Eine triviale Transformation ist beispielsweise die charakteristische Funktion des Zielkonzepts.