Hans U. Simon
Malte Darnstädt

Bochum, den 21.04.2014
Abgabe bis zum 28.04.2014

Übungen zur Vorlesung
Theorie des Maschinellen Lernens
Sommer 2014

Übungsblatt 02

Aufgabe 2.1

Führe den Algorithmus Greedy zum Lernen von Monomen mit wenigen Literalen mit folgender Eingabe von Hand aus. Schreibe auf, wie sich die Mengen L_0 (die Menge der Literale, die in die Hypothese aufgenommen werden) und U (die Menge der unüberdeckten negativen Beispiele aus T) im Verlauf des Algorithmus verändern. Findet der Algorithmus das kürzeste konsistente Monom?

Sei $n = 5$. Die Beispielmenge T ist der Einfachheit halber durchnummeriert:

0	1	2	3	4
$(01110, +)$	$(11001, -)$	$(00010, -)$	$(11010, -)$	$(00111, -)$

5	6	7	8
$(00001, -)$	$(11000, -)$	$(00110, -)$	$(11110, -)$

Aufgabe 2.2

Sei $X_n = \{0, 1, \ldots, n-1\}$. Wir betrachten die Konzeptklasse $P_{k,n}$, die durch das Vorzeichen von Polynomen von Grad k gebildet wird; das heißt die Elemente von $P_{k,n}$ haben die folgende Form:

$$c(x) = \begin{cases} 1 & \text{falls } \sum_{i=0}^{k} a_i x^i \geq 0 \\ 0 & \text{sonst} \end{cases}$$

mit $a_i \in \mathbb{R}$.

a) Wie viele funktionell verschiedene Konzepte sind in $P_{k,n}$ enthalten?

b) Sei k fest. Gib einen effizienten PAC-Lerner für $(P_{k,n})_{n \geq 1}$ an.

Aufgabe 2.3

Von folgendem Problem ist bekannt, dass es nicht effizient lösbar ist (falls $RP \neq NP$):

Problem: Set-Splitting
Eingabe: Eine Menge $S = \{x_1, \ldots, x_n\}$ und Teilmengen $A_1, \ldots, A_l$ von S.
Frage: Gibt es Teilmengen S_1, S_2 von S, so dass $S = S_1 \cup S_2$ gilt und kein A_i vollständig in S_1 oder S_2 liegt?

Zeige durch eine Reduktion von Set-Splitting auf das Konsistenzproblem für 2-Term-DNF, dass die Klasse 2-Term-DNF nicht effizient proper PAC-lernbar ist (falls $RP \neq NP$).

Aufgabe 2.4

Ein randomisierter Lerner hat – neben dem üblichen Beispielorakel – Zugriff auf ein Münzorakel, das mit Wahrscheinlichkeit 1/2 entweder mit *Kopf* oder *Zahl* antwortet.

Sei h_0 die konstante Null- und h_1 die konstante Eins-Funktion. Zeige: Wenn eine Konzeptklasse $\mathcal{C}$ mit Hypothesenklasse $\mathcal{H}$ von einem randomisierten Lerner effizient PAC-lernbar ist, dann ist $\mathcal{C}$ mit $\mathcal{H} \cup \{h_0, h_1\}$ von einem deterministischen Lerner effizient PAC-lernbar.

Hinweise: Nimm an, dass $\epsilon < 1/4$ gilt. Überlege dir, wie man mit Hilfe von Beispielorakel-Anfragen eine zufällige Münze simulieren kann. Verwende außerdem die Aussage von Hausaufgabe 1.4. und Präsenzaufgabe 1.4.