Hans U. Simon
Christoph Ries

Bochum, den 23. Mai 2016
Abgabe bis zum 31.05.2016

# Übungen zur Vorlesung
# **Theorie des maschinellen Lernens**
## Sommer 16
## Übungsblatt 05

**Aufgabe 5.1** (4 Punkte)

In der Vorlesung wurde INDEPENDENT SET auf die Entscheidungsvariante des Minimum Disagreement Problems bzgl. monotonen, boolschen Konjunktionen reduziert. Dabei wurde eine Instanz des INDEPENDENT SET-Problems $(G = ([n], E), k)$ wie folgt in eine Trainingsmenge $S$ für das Entscheidungsproblem transformiert. Für alle $i = 1, ..., n$ ist $(\bar{0}_i, 1)$ und für alle $\{i, j\} \in E$ ist $(\bar{0}_{i,j}, 0)$ in $S$. Die zugehörige Fehlerschranke wurde auf $n - k$ gesetzt.

a) Gegeben sei eine Eingabe $(G, k)$ des INDEPENDENT SET-Problems mit $G = ([5], E)$, $E = \{\{1, 3\}, \{1, 4\}, \{1, 5\}, \{2, 3\}, \{3, 4\}, \{3, 5\}\}$ und $k = 3$. Führe obige Transformation beispielhaft bzgl. $(G, k)$ durch.

b) Zeige, dass sich die obige Transformation ebenfalls eignet, um folgende Behauptung zu verifizieren.
Behauptung: Es existiert genau dann ein Independent Set $U \subseteq [n]$ in $G$ mit $|U| \leq k$, wenn eine boolsche Konjunktion existiert, die auf $S$ höchstens $n-k$ Fehler macht.

**Aufgabe 5.2** (4 Punkte)

In der Vorlesung wurde 2-FÄRBBARKEIT VON HYPERGRAPHEN auf die Entscheidungsvariante des Konsistenzproblems bzgl. monotonen 2-Term DNF reduziert. Dabei wurde eine Instanz des 2-FÄRBBARKEIT VON HYPERGRAPHEN-Problems $([n], \{T_1, ..., T_m\})$ mit $T_j \subseteq [n]$ wie folgt in eine Trainingsmenge $S$ für das Entscheidungsproblem transformiert. Für alle $i = 1, ..., n$ ist $(\bar{0}_i, 1)$ und für alle $j = 1, ..., m$ ist $(\bar{0}_{T_j}, 0)$ in $S$.

a) Gegeben sei eine Eingabe $([5], \{T_1, ..., T_4\})$ des 2-FÄRBBARKEIT VON HYPERGRAPHEN-Problems mit $T_1 = \{1, 4\}, T_2 = \{1, 3, 5\}, T_3 = \{2, 3\}$ und $T_4 = \{2, 5\}$. Führe obige Transformation beispielhaft bzgl. $([5], \{T_1, ..., T_4\})$ durch.

b) Zeige, dass sich die obige Transformation ebenfalls eignet, um folgende Behauptung zu verifizieren.
Behauptung: Es existiert genau dann eine 2-Färbung für $([n], \{T_1, ..., T_m\})$, wenn eine 2-Term DNF existiert, die auf $S$ fehlerfrei ist.

— Bitte wenden! —

**Aufgabe 5.3** (4 Punkte)
Seien $\mathcal{H}_1 := \{h_u \mid u \in \mathbb{R}^d\}$ und $\mathcal{H}_2 := \{h_{v,w} \mid v, w \in \mathbb{R}^d\}$ zwei Hypothesenklassen, wobei

$$h_u(x) := \begin{cases} 1 & \text{, falls } \langle u, x\rangle > 0 \\ 0 & \text{, sonst} \end{cases}$$

und

$$h_{v,w}(x) := \begin{cases} 1 & \text{, falls } (\langle v, x\rangle > 0 \land \langle w, x\rangle < 0) \lor (\langle v, x\rangle < 0 \land \langle w, x\rangle > 0) \\ 0 & \text{, sonst} \end{cases}.$$

Zeige, dass sich $\mathcal{H}_2$ auf $\mathcal{H}_1$ PAC-reduzieren lässt.

**Aufgabe 5.4** (4 Punkte)
Sei $\mathcal{H} := \{I_{a,b} \mid a, b \in \mathbb{R} \land a \leq b\}$ mit

$$I_{a,b}(x) = \begin{cases} 1 & \text{, falls } a \leq x \leq b \\ 0 & \text{, sonst} \end{cases}.$$

Schlage eine Implementierung der ERM-Regel vor, die „möglichst effizient“ ist.

*Hinweis*: Es existiert ein Algorithmus, der bei einer Trainingsgröße $m$ die Laufzeit $\mathcal{O}(m \log(m))$ hat.