Hans U. Simon
Christoph Ries

Bochum, den 26. April 2016
Abgabe bis zum 03.05.2016

Übungen zur Vorlesung

Theorie des maschinellen Lernens

Sommer 16

Übungsblatt 02

Aufgabe 2.1 (4 Punkte)
Wir betrachten die Klasse aller achsenparallelen Rechtecke

$$\mathcal{H}^2_{rec} := \{h_{(a_1,b_1,a_2,b_2)} : a_1 \le b_1 \wedge a_2 \le b_2\}.$$

Eine Hypothese aus $\mathcal{H}^2_{rec}$ hat die Form

$$h_{(a_1,b_1,a_2,b_2)}(x_1,x_2) = \begin{cases} 1 & \text{, falls } a_1 \le x_1 \le b_1 \wedge a_2 \le x_2 \le b_2 \\ 0 & \text{, sonst} \end{cases} .$$

Wir betrachten im Folgenden den realisierbaren Fall.

a) Sei A ein Algorithmus. A bestimme das kleinste Rechteck, welches alle positiven Punkte der Trainingsmenge umschließt, und gebe die zugehörige Hypothese aus. Zeige, dass A ein ERM-Algorithmus ist.

b) Skizziere analog zu den Halbintervallen aus der Vorlesung einen Beweis, dass A bei Eingabe einer Trainingsmenge S mit

$$|S| \ge \frac{4\log(4/\delta)}{\varepsilon}$$

mit einer Wahrscheinlichkeit von mindestens $1-\delta$ eine Hypothese ausgibt deren Fehler höchstens ε ist.

c) Wieviele Beispiele braucht ein entsprechender Algorithmus für achsenparallele d-dimensionale Boxen (ein Rechteck ist eine zwei-dimensionale Box) in $\mathbb{R}^d$? Beweise deine Behauptung.

— Bitte wenden! —

Aufgabe 2.2 (4 Punkte)

Seien $D_1, ..., D_m$ Verteilungen über X. Sei $\mathcal{H}$ eine endliche Hypothesenklasse mit Hypothesen von X nach $\{0,1\}$. Sei $f \in \mathcal{H}$. Es wird eine Trainingsmenge $S = \{x_1, ..., x_m\}$ der Größe m unabhängig, aber nicht gleichverteilt gezogen, d.h. $\forall i$: $x_i \sim D_i$ und x_i wird mit f markiert. Des Weiteren sei

$$\bar{D} := \frac{1}{m} \sum_{i=1}^{m} D_i.$$

Sei $\varepsilon \in (0,1)$ beliebig, aber fest. Zeige, dass

$$Pr[\exists h \in \mathcal{H} : L_{\bar{D},f}(h) > \varepsilon \wedge L_{(S,f)}(h) = 0] \leq |\mathcal{H}| \cdot e^{-\varepsilon m}.$$

Aufgabe 2.3 (4 Punkte)

Sei $|X| \geq km$ für $k \geq 2$. Zeige, dass die untere Grenze $1/4$ im No-Free-Lunch-Theorem durch $\frac{k-1}{2k}$ ersetzt werden kann, d.h.:
Sei A ein Algorithmus zur binären Klassifizierung. Sei $m \leq |X|/k$ die Größe der Trainingsmenge. Zeige, dass dann eine Verteilung D über $X \times \{0,1\}$ existiert, sodass gilt:

(i) Es existiert eine Funktion $f : X \to \{0,1\}$ mit $L_D(f) = 0$.

(ii) Mit Wahrscheinlichkeit mindestens $\frac{k-1}{3k+1}$ (über $S \sim D^m$) gilt

$$L_D(A(S)) \geq \frac{k-1}{4k}.$$

Aufgabe 2.4 (4 Punkte)

Sei $X = \mathbb{R}$. Die Hypothesenklasse sei durch

$$\mathcal{H} := \{x \mapsto \lceil 0.5 \cdot \sin(\theta x) \rceil : \theta \in \mathbb{R}\}$$

gegeben. Zeige, dass $VCdim(\mathcal{H}) = \infty$.

TIPP: *Eine Möglichkeit besteht darin, folgende Aussage mit Beweis zu verwenden. Sei* $0.x_1x_2x_3...$ *die binäre Darstellung von* $x \in (0,1)$, *dann gilt für alle* $m \in \mathbb{N}$

$$\lceil 0.5 \cdot \sin(2^m \pi x) \rceil = 1 - x_m$$

vorausgesetzt, dass ein $k \geq m$ *existiert mit* $x_k = 1$.