Hans U. Simon
Malte Darnstädt

Bochum, den 23.04.2013
Abgabe bis zum 01.05.2013

Übungen zur Vorlesung

Theorie des Maschinellen Lernens

Sommer 2013

Übungsblatt 02

Aufgabe 2.1

Zeige folgende Lemmata für Aufgabe 2.2:

a) (eine Variante von Lemma 4.4 aus dem Buch)

Sei $Q = \{z \in Z^m : \exists h \in H \text{ mit } \hat{\mathrm{er}}_z(h) = 0 \wedge \mathrm{er}_P(h) > \epsilon\}$
und $R = \{(z, z') \in Z^{2m} : \exists h \in H \text{ mit } \hat{\mathrm{er}}_z(h) = 0 \wedge \hat{\mathrm{er}}_{z'}(h) > \epsilon/2\}$.
Dann gilt $P^m(Q) \leq 2 \cdot P^{2m}(R)$ für alle $m \geq \frac{2}{\epsilon}$.

b) (eine Variante von Lemma 4.6 aus dem Buch)

Sei Γ_m die Menge der Permutationen auf $\{1, \ldots, 2m\}$, für die für jedes $1 \leq i \leq m$ entweder $\sigma(i) = m+i, \sigma(m+i) = i$ oder $\sigma(i) = i, \sigma(m+i) = m+i$ gilt. Sei $(z, z') \in Z^{2m}$ fest und $\sigma(z, z')$ die entsprechend permutierte Stichprobe.

Dann gilt für eine aus Γ_m gleichverteilt gezogene Permutation σ:
$\Pr_\sigma(\sigma(z, z') \in R) \leq \Pi_H(2m) \cdot 2^{-m\epsilon/2}$

Hinweis: Mit $z = ((x_1, y_1), \ldots, (x_m, y_m))$, $z' = ((x_{m+1}, y_{m+1}), \ldots, (x_{2m}, y_{2m}))$ gilt

$$\Pr_\sigma\Big(\sigma(z, z') \in R\Big) = \Pr_\sigma\Big(\exists h \in H : \sum_{i=1}^{m} 1(h(x_{\sigma(i)}) \neq y_{\sigma(i)}) = 0 \text{ und } \sum_{i=1}^{m} 1(h(x_{\sigma(m+i)}) \neq y_{\sigma(m+i)}) > \frac{\epsilon m}{2}\Big)$$

Aufgabe 2.2

Beweise die obere Schranke im Restricted Model aus der Vorlesung (Theorem 4.8 im Buch), d.h. zeige dass für jede Klasse H mit VC-Dimension $d \geq 1$ und jeden Algorithmus L, der eine konsistente Hypothese ausgibt, folgende Abschätzungen für den Schätzfehler und die Label-Komplexität gelten:

$$\epsilon_L(m, \delta) \leq \frac{2}{m}\left(d \log_2\left(\frac{2em}{d}\right) + \log_2\left(\frac{2}{\delta}\right)\right)$$
$$m_L(\epsilon, \delta) \leq \frac{4}{\epsilon}\left(d \log_2\left(\frac{12}{\epsilon}\right) + \log_2\left(\frac{2}{\delta}\right)\right)$$

Hinweis: Verwende die Aussagen von Aufgabe 2.1 und folge dem Beweis der Schranke für das allgemeine Modell.

Aufgabe 2.3

Sei $X_n = \{0,1\}^n$ und H_n die Menge der *boolean Perceptrons*, also die Menge der Abbildungen $f : X_n \to \{0,1\}$ die sich über

$$f(x) = \mathrm{sgn}\Big(\sum_{i=1}^{n} w_i x_i - \theta\Big)$$

mit $w \in \mathbb{R}^n$ und $\theta \in \mathbb{R}$ darstellen lassen.

Zeige, dass für jedes $f \in H_n$ ein passendes $w \in \mathbb{R}^n$ existiert, so dass $2^n + 1$ verschiedene Elemente aus H_n alleine durch Änderungen von θ gebildet werden können.

Aufgabe 2.4

Sei H_n wie in Aufgabe 2.3 definiert. Zeige:

$$|H_n| \geq 2^{(n^2-n)/2}$$

Hinweis: Verwende die Aussage von Aufgabe 2.3.