

Algorithmische Grundlagen des Maschinellen Lernens
Sommersemester 2022
Übungsblatt 4

Aufgabe 1: (3 Punkte)

Es seien S und S' zwei Mengen von jeweils m Samples unabhängig gezogen aus der gleichen Verteilung \mathcal{D} . Zeigen Sie mittels geschickter Anwendung der Hoeffding-Ungleichung, dass für jede Hypothese h gilt

$$\Pr[|\text{err}_S(h) - \text{err}_{S'}(h)| \geq \gamma] \leq 2 \exp(-m\gamma^2/2).$$

Aufgabe 2: (3 Punkte)

Gegeben eine Hypothesenklasse \mathcal{H} über der Grundmenge \mathcal{X} . Zeigen Sie, ist \mathcal{H} PAC-lernbar im agnostischen Sinn, so ist \mathcal{H} auch PAC-lernbar im realisierbaren Sinn.

Aufgabe 3: (4 Punkte)

Angenommen, wir haben einen Lernalgorithmus, der den Trainingsfehler nur approximativ minimiert. Das heißt, für alle Trainingsmengen S gilt $\text{err}_S(h_S) \leq \min_{h \in \mathcal{H}} \text{err}_S(h) + \gamma$ für ein festes $\gamma > 0$.

Zeigen Sie, dass für jede Wahl von $\epsilon > 0$, $\delta > 0$, wenn S eine m -elementige Trainingsmenge aus einer Datenpunkt-/Label-Verteilung \mathcal{D} gezogen ist mit

$$m \geq \frac{2}{\epsilon^2} \ln \left(\frac{2|\mathcal{H}|}{\delta} \right) ,$$

dann gilt $\text{err}_{\mathcal{D}}(h_S) \leq \min_{h' \in \mathcal{H}} \text{err}_{\mathcal{D}}(h') + \epsilon + \gamma$ mit Wahrscheinlichkeit mindestens $1 - \delta$.

Aufgabe 4: (5+5 Punkte)

Bestimmen Sie jeweils die VC-Dimension der folgenden Mengensysteme. Beweisen Sie Ihre Antwort.

- \mathcal{R} besteht aus Halbräumen der Form $r = \{ x \in \mathbb{R}^d \mid \langle a, x \rangle \geq 1 \}$ und Halbräumen der Form $r = \{ x \in \mathbb{R}^d \mid \langle a, x \rangle \leq 1 \}$ mit $a \in \mathbb{R}^d$.
- \mathcal{R} besteht aus Halbräumen der Form $r = \{ x \in \mathbb{R}^d \mid \langle a, x \rangle \geq 0 \}$ mit $a \in \mathbb{R}^d$.

Tipp: Nutzen Sie aus, dass $d + 1$ Vektoren im \mathbb{R}^d immer linear abhängig sind.