Hans U. Simon
Christoph Ries

Bochum, den 13. Juni 2016
Abgabe bis zum 21.06.2016

Übungen zur Vorlesung

Theorie des maschinellen Lernens

Sommer 16

Übungsblatt 08

Aufgabe 8.1 (4 Punkte)
Sei A ein Lernalgorithmus. Es existiere ein $\delta_0 \in (0,1)$ und eine Funktion $m_{\mathcal{H}} : (0,1) \to \mathbb{N}$, sodass für alle $\varepsilon \in (0,1)$, alle Verteilungen D und alle $m \geq m_{\mathcal{H}}(\varepsilon)$, gilt

$$Pr_{S \sim D^m}[L_D(A(S)) \leq \min_{h \in \mathcal{H}} L_D(h) + \varepsilon] \geq 1 - \delta_0.$$

Schlage ein Verfahren vor, welches A verwendet und $\mathcal{H}$ agnostisch PAC-lernt mit

$$m_{\mathcal{H}}(\varepsilon, \delta) \leq k \cdot m_{\mathcal{H}}(\varepsilon) + \left\lceil \frac{2\log(4k/\delta)}{\varepsilon^2} \right\rceil,$$

wobei $k = \lceil \log(\delta)/\log(\delta_0) \rceil$.

Hinweis Teile dazu die Trainingsmenge in $k+1$ Bereiche, sodass die ersten k aus jeweils $m_{\mathcal{H}}(\varepsilon)$ Beispielen bestehen. Bestimme mit Hilfe von A für jeden dieser k Bereiche eine Hypothese. Zeige, dass $Pr[L_D(A(S)) > \min_{h \in \mathcal{H}} L_D(h) + \varepsilon] \leq \delta_0^k \leq \delta/2$ für jeden dieser Bereich. Nutze abschließend den letzten Bereich, um eine dieser k Hypothesen auszuwählen.

Aufgabe 8.2 (4 Punkte)

a) Betrachte die Trainingsmenge $S = \{((0,0),1),((1,0),-1),((0,1),-1),((1,1),1)\}$. Lasse AdaBoost mit Decision Stumps als Basislerner auf S laufen. Was ist der kleinste Fehler den AdaBoost auf S realisieren kann? Begründe.
Beachte, dass der Basislerner in diesem Fall kein γ-schwacher Lerner ist.

b) Welche Hypothese gibt AdaBoost mit Decision Stumps als Basislerner nach zwei Runden auf der Trainingsmenge

$$S = \{(1,1),(2,1),(3,1),(4,-1),(5,-1),(6,-1),(7,1)\}$$

aus.

c) AdaBoost laufe auf einer Trainingsmenge der Größe m. In jeder Runde sei der gewichtete Fehler ε_t der t-ten schwachen Hypothese höchsten $1/2 - \gamma$, wobei $0 < \gamma < 1/2$. Nach wievielen Iterationen $T(m,\gamma) \in \mathbb{N}$ wird die zusammengesetzte Hypothese konsistent mit den m Trainingsbeispielen sein?

— Bitte wenden! —

Aufgabe 8.3 (4 Punkte)

Betrachte den Algorithmus AdaBoost. Zeige, dass der Fehler von h_t bezüglich der Verteilung $D^{(t+1)}$ exakt 1/2 ist, d.h. zeige, dass für alle $t \in [T]$

$$\sum_{i=1}^{m} D_i^{(t+1)} \mathbb{1}_{[y_i \neq h_t(x_i)]} = 1/2.$$

Aufgabe 8.4 (4 Punkte)

Bei der k-fachen Kreuz-Validierung wird eine Trainingsmenge S der Größe m in k Teile $S_1, ..., S_k$ der Größe m/k zerlegt. Der Fall $k = m$ heißt Leave-one-out (LOO).
Gegeben sei eine Grundmenge $\mathcal{X}$ mit einer Verteilung D. Eine Instanz $x \in \mathcal{X}$ wird zufällig mit $y = 1$ bzw. $y = 0$ markiert - es soll dabei gelten $\Pr[y = 1 \mid x] = \Pr[y = 0 \mid x] = 1/2$. Sei $S = \{(x_1, y_1), ..., (x_m, y_m)\}$ eine Trainingsmenge. Ein Lernalgorithmus A liefere bei Eingabe von S die Hypothese h, wobei

$$h(x) = \begin{cases} 1 & \text{, falls } \sum_{i=1}^{|S|} y_i = 1 \mod 2 \\ 0 & \text{, sonst} \end{cases}.$$

Gemäß LOO sei $S_i = \{(x_i, y_i)\}$ für alle $1 \leq i \leq k = m$. Zeige, dass der mittlere Fehler

$$\frac{1}{k} \sum_{i=1}^{k} L_{S_i}(A(S \setminus S_i))$$

entweder 0 oder 1 ist, obwohl der tatsächliche Fehler (bei komplett zufälligen Labels) 1/2 ist.