Hans U. Simon
Malte Darnstädt

Bochum, den 09.05.2014
Abgabe bis zum 19.05.2014

Übungen zur Vorlesung

Theorie des Maschinellen Lernens

Sommer 2014

Übungsblatt 05

Beachte: Auf diesem Aufgabenblatt arbeiten wir im Boosting-Framework, in dem Konzepte nicht mehr auf $\{0, 1\}$, sondern auf $\{-1, 1\}$ abbilden.

Aufgabe 5.1

Sei $X = \mathbb{R}$ und sei $\mathcal{H}$ die Klasse der Halbintervalle, die aus folgenden Hypothesen $h_{t,b}$ besteht:

$$h_{\theta,b}(x) = \begin{cases} b, & \text{falls } x \geq \theta \\ -b, & \text{falls } x < \theta \end{cases}$$

für $\theta \in \mathbb{R}$ und $b \in \{-1, 1\}$.

Sei folgende Menge S gegeben:

$$S = \{(1, +1),\ (2, -1),\ (3, +1),\ (4, -1)\}$$

Führe drei Iterationen von AdaBoost mit der Basisklasse der Halbintervalle $\mathcal{H}$ aus. Wähle in jedem Schritt ein h_t mit minimalen Fehler ϵ_t.

Für die Abgabe reicht es, folgendes zu notieren: D_t, h_t, ϵ_t, α_t (für $1 \leq t \leq 3$) und die endgültige Hypothese $H(x)$. Reelle Zahlen können dabei auf drei Nachkommastellen gerundet werden. Gib außerdem die Anzahl der Fehler von H auf S an.

Aufgabe 5.2

Zeige bei AdaBoost, dass der Fehler von h_t unter Verteilung D_{t+1} genau $\frac{1}{2}$ ist.

Zeige also:

$$\Pr_{i \sim D_{t+1}}(h_t(x_i) \neq y_i) = \frac{1}{2}$$

Aufgabe 5.3

Betrachte folgende randomisierte Variante von AdaBoost: Die zusammengesetzte Hypothese H wird durch einen Klassifizierer $\tilde{H}$ ersetzt, der mit einer Wahrscheinlichkeit von $\frac{e^{F(x)}}{e^{F(x)}+e^{-F(x)}}$ für den Punkt x das Label $+1$ vorhersagt (ansonsten sagt er -1 voraus). Dabei ist $F(x) = \sum_{t=1}^{T} \alpha_t h_t(x)$.

Zeige, dass die obere Schranke für den Fehler auf den Trainingsdaten aus der Vorlesung für $\tilde{H}$ um den Faktor 2 verbessert werden kann. Das heißt, zeige:

$$\Pr_{i \sim D_1, \tilde{H}}(\tilde{H}(x_i) \neq y_i) \leq \frac{1}{2} \prod_{t=1}^{T} \sqrt{1 - 4\gamma_t^2}$$

Aufgabe 5.4

Sei $X = \{0,1\}^n$ und sei S eine Stichprobe mit Punkten aus $X \times \{-1, 1\}$, die von einer k-Term-DNF gelabelt worden sind. Weise nach, dass die Basisklasse der Monome auf S einen Margin von mindestens $\frac{1}{2k-1}$ besitzt.

Zeige also: Für jede k-Term-DNF $c : X \to \{-1, 1\}$ existiert eine gewichtete Summe

$$h(x) = \sum_{i=1}^{l} \alpha_i M_i(x) \quad ,$$

wobei die $M_i : X \to \{-1, 1\}$ durch Monome gegeben sind und $\alpha_i \geq 0$, $\sum_{i=1}^{l} \alpha_i = 1$, mit der Eigenschaft

$$c(x) \cdot h(x) \geq \frac{1}{2k-1} \qquad \text{für alle } x \in X.$$

Hinweis: Die Basisklasse der Monome enthält mit dem leeren Monom die konstante-1-Funktion.