Hans U. Simon
Christoph Ries

Bochum, den 20. Juni 2017
Abgabe bis zum 27.06.2017

Übungen zur Vorlesung

Theorie des maschinellen Lernens

Sommer 17

Übungsblatt 08

Aufgabe 8.1 (4 Punkte)
Sei $h_w : \mathbb{R}^3 \to [0,1]$ mit $w \in \mathbb{R}^3$. Es gelte

$$h_{(w_1,w_2,w_3)}(x_1,x_2,x_3) := \frac{1}{1+e^{-(w_1x_1+w_2x_2+w_3x_3)}}.$$

Gib für die Trainingsmenge

$$S = \{((-1,2,1),-1),((2,6,1),1),((-2,5,1),1),$$
$$((3,0,1),1),((1,0,1),-1),((-3,-1,1),-1)\}$$

ein w an, sodass alle Punkte aus S durch $\mathrm{sign}(\langle w,x\rangle)$ korrekt klassifiziert werden. Bestimme anschließend die Wahrscheinlichkeiten, die durch deine Wahl von w jedem Punkt aus S durch h_w zugeordnet werden (vgl. Lemma 9.18 im Skriptteil „Lineare Voraussagefunktionen“ unter Materialien auf der Homepage).

Aufgabe 8.2 (4 Punkte)
Gegeben sei eine Trainingsmenge $S = \{(x_i,y_i)\}_{i=1}^m \in (\mathcal{X} \times \{-1,1\})^m$, Hypothesen $g_1,...,g_T$: $\mathcal{X} \to \{-1,1\}$, sowie $a_1,...,a_T \in [0,1]$ mit $\sum_{i=1}^T a_i = 1$. Es gelte für ein $\theta > 0$, dass

$$y_i \sum_{j=1}^{T} a_j g_j(x_i) \geq \theta$$

für alle $1 \leq i \leq m$.

a) Zeige, dass sich die y_i für alle $1 \leq i \leq m$ durch ein gewichtetes Majoritätskriterium der $g_1,...,g_T$ bestimmen lassen.

b) Sei D eine Verteilung über $[m]$. Zeige, dass mit obigen Voraussetzungen und $\theta = 2\gamma$ eine Hypothese g_j existiert mit

$$Pr_{i\sim D}[y_i \neq g_j(x_i)] \leq \frac{1}{2} - \gamma.$$

— Bitte wenden! —

Aufgabe 8.3 (4 Punkte)

a) Betrachte die Trainingsmenge $S = \{((0,0),1),((1,0),-1),((0,1),-1),((1,1),1)\}$. Lasse AdaBoost mit Entscheidungsstümpfen 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 Entscheidungsstümpfen als Basislerner nach zwei Runden auf der Trainingsmenge

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

aus.

Aufgabe 8.4 (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 $t \in [T]$

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