Hans U. Simon
Christoph Ries

Bochum, den 5. Juni 2018
Abgabe bis zum 12.06.2018

Übungen zur Vorlesung

Theorie des maschinellen Lernens

Sommer 18

Übungsblatt 06

Aufgabe 6.1 (4 Punkte)
Wir betrachten eine Variante des Perzeptron-Algorithmus. Wir ersetzen dazu den Aktualisierungsschritt $w^{(t+1)} = w^{(t)} + y_i x_i$ durch

$$w^{(t+1)} = w^{(t)} + \eta y_i x_i$$

für ein $\eta > 0$.
Zeige, dass diese Variante des Perzeptrons dieselbe obere Schranke für die Anzahl der Iterationen wie die Standardvariante hat.

Aufgabe 6.2 (4 Punkte)
Wir betrachten nun lineare Regression mit der Verlustfunktion $l(h,(x,y)) = |h(x) - y|$. Für eine Trainingsmenge $\{(x_i, y_i)\}_{i=1}^m$, wobei $x_i \in \mathbb{R}^d$ und $y_i \in \mathbb{R}$, soll gemäß der ERM-Regel das w gefunden werden, welches den Ausdruck

$$\sum_{i=1}^{m} |\langle w, x_i \rangle - y_i|$$

minimiert.
Zeige, dass sich obige Problemstellung in ein lineares Programm umwandeln lässt.

— Bitte wenden! —

Aufgabe 6.3 (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

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

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 7.18 im Skript auf der Homepage unter Materialien).

Aufgabe 6.4 (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

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