Anne Driemel
Thomas Kesselheim
Xianghui Zhong

15. Juni 2020
Abgabe bis: 22. Juni 2020, 12 Uhr

Algorithmische Grundlagen des Maschinellen Lernens

Sommersemester 2020

Übungsblatt 8

Aufgabe 1: (5+2 Punkte)
Wir betrachten $X = \mathbb{R}$, $Y = \{-1, 1\}$ und $\mathcal{H}$ als die Klasse aller Schwellenwertfunktionen ($h_a(x) = 1$, falls $x \geq a$, -1 sonst). Als Loss-Funktionen betrachten wir den 0/1 Loss, definiert über $\ell^{0-1}(h, z) = 0$ wenn z durch h korrekt klassifiziert wird, 1 sonst.

(a) Sei $\mathcal{A}$ ein Algorithmus, der den Trainingsfehler L_S^{0-1} minimiert. Zeigen Sie, dass $\mathcal{A}$ nicht universell austauschstabil ist. Das heißt, konstruieren Sie für alle $m \in \mathbb{N}$ ein S, i und z', sodass $\ell(h_{S^i}, z_i) - \ell(h_S, z_i) \geq 1$.
Tipp: Dies tritt bereits in realisierbaren Instanzen auf.

(b) Begründen Sie kurz mithilfe von Satz 5.1, dass der erwartete Verallgemeinerungsfehler trotzdem für $m \to \infty$ verschwindet.

Aufgabe 2: (4 Punkte)
Wir betrachten eine Menge von Datenpunkte in $\mathbb{R}$ mit Labels $S = \{(x_1, y_1), \ldots, (x_m, y_m)\} \subseteq \mathbb{R} \times \{-1, 1\}$. Konstruieren Sie eine Hypothese h^* als Linearkombination von Schwellenwertfunktionen, sodass $h(x_i) = y_i$ für alle i. Sie dürfen dabei annehmen, dass S keine zwei Punkte enthält mit $x_i = x_j$ und $y_i \neq y_j$. Es sind auch negative Koeffizienten in der Linearkombination erlaubt.

Aufgabe 3: (4 Punkte)
Zeigen Sie, dass für alle Schritte t in AdaBoost gilt

$$\sum_{i: h_t(x_i) \neq y_i} p_i^{(t+1)} = \frac{1}{2} \ .$$

Diese Gleichung gibt auch eine schöne intuitive Erklärung für die Wahl von $p^{(t+1)}$: Es handelt sich um eine Wahrscheinlichkeitsverteilung, sodass h_t nicht besser ist als zufälliges Raten.

Aufgabe 4: (5 Punkte)
Sei S eine Menge von m Datenpunkt-/Label-Paaren $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m)$, wobei $\mathbf{x}_i = (x_{i,1}, x_{i,2}) \in \mathbb{R}^2$ und

$$y_i = \begin{cases} +1 & \text{falls } x_{i,1} \geq 0 \text{ und } x_{i,2} \geq 0 \\ -1 & \text{sonst} \end{cases} \ .$$

Zeigen Sie, dass es für jede solche Menge S einen linearen Klassifikator $h_{\mathbf{w},u}$ ($\mathbf{w} \in \mathbb{R}^2, u \in \mathbb{R}$, $h_{\mathbf{w},u}(x) = 1$, falls $\langle \mathbf{w}, x \rangle \geq u$, -1 sonst) gibt mit $\mathrm{err}_S(h_{\mathbf{w},u}) \leq \frac{1}{3}$. Beachten Sie, dass für $\mathbf{w}$ auch der Nullvektor zugelassen ist.