Universität Salzburg — Florian Graf

Machine Learning

Übungsblatt **10** — 20 Punkte

Aufgabe 1. *Adaptives Boosting I* — 14 P.

Gegeben seien n Daten-Label-Paare $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\} \subset \mathbb{R}^d \times \{-1, 1\}$. Wir trainieren mittels adaptivem Boosting (AdaBoost) ein additives Entscheidungsmodell

$$f(\mathbf{x}; \boldsymbol{\theta}) = \sum_{m=1}^{M} \beta_m F_m(\mathbf{x}; \boldsymbol{\theta})$$

mit Basismodellen F_m und Gewichten $\beta_m > 0$.
Es sei

$$\epsilon_m = \frac{1}{Z_m} \sum_{i=1}^{n} w_{i,m} \mathbb{1}_{y_i \neq F_m(\mathbf{x}_i)}$$

der gewichtete Fehler des m-ten Modells, mit Gewichten $w_{i,m}$ und Normierungskonstante $Z_m = \sum_{i=1}^n w_{i,m}$. Im AdaBoost Verfahren sind die Modellgewichte definiert als $\beta_m = \frac{1}{2} \log\left(\frac{1-\epsilon_m}{\epsilon_m}\right)$, und die jeweiligen Fehlergewichte sind definiert als

$$w_{i,m+1} = \begin{cases} w_{i,m} \exp(-\beta_m y_i F_m(\mathbf{x}_i)) & m > 0 \\ 1 & m = 0 \end{cases} .$$

In dieser Aufgabe zeigen Sie, dass die Genauigkeit des geboosteten Modells f auf den Trainingsdaten exponentiell mit der Anzahl der Basismodelle abnimmt, solange jedes Basismodell einen Fehler von weniger als 50% macht, also dass für $\gamma = \min_m(1/2 - \epsilon_m)$ gilt, dass

$$\mathrm{err}(f) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{y_i \neq f(\mathbf{x}_i)} \leq \exp(-2\gamma^2 M) \quad .$$

(a) Beweisen Sie zunächst die Identität $\mathbb{1}_{t \leq 0} \leq \exp(-t)$ für $t \in \mathbb{R}$ und folgern Sie daraus, dass $\mathbb{1}_{y_i \neq f(\mathbf{x}_i)} \leq e^{-y_i f(\mathbf{x}_i)}$.

(b) Folgern Sie aus der Definition der Gewichte $w_{i,m}$ dass $w_{i,M} = e^{-y_i f(\mathbf{x}_i)}$.

(c) Zeigen Sie außerdem, dass die Normierungskonstanten Z_m die folgende Rekursion erfüllen

$$Z_{m+1} = \begin{cases} 2Z_m \sqrt{\epsilon_m (1 - \epsilon_m)} & m > 0 \\ n & m = 0 \end{cases} .$$

Zwischenergebnis zur Kontrolle: $Z_{m+1} = Z_m \left((1 - \epsilon_m)e^{-\beta_m} + \epsilon_m e^{\beta_m}\right)$ für $m > 0$.

(d) Folgern Sie aus (a) und (b), dass $\mathrm{err}(f) \leq \frac{Z_M}{n}$, und dann aus (c), dass $\mathrm{err}(f) \leq \prod_{m=1}^{M} \sqrt{4\epsilon_m(1 - \epsilon_m)}$.

(e) Zeigen Sie mithilfe der Ungleichung $1 - t \leq e^{-t}$, dass $4\epsilon_m(1 - \epsilon_m) \leq \exp^{-(2\epsilon_m - 1)^2}$.

(f) Zeigen Sie schließlich, dass $\mathrm{err}(f) \leq \exp -2\gamma^2 M$.

Aufgabe 2. *Adaptives Boosting II* — 6 P.

Wir nutzen die gleiche Notation wie in Augabe 1.

(a) Es sei $k \in (0, 1)$ fest und $\phi : \mathbb{R} \to \mathbb{R}, t \mapsto (1 - k)e^{-t} + ke^t$. Berechnen Sie den Minimierer $\hat{t}$ der Funktion ϕ.

(b) Interpretieren Sie (a) hinsichtlich des AdaBoost Verfahrens.

(c) Zeigen Sie, dass im AdaBoost Verfahren jedes Basismodell F_{m+1} die korrekt und falsch klassifizierten Daten des Modells F_m insgesamt gleich gewichtet, d.h, dass

$$\frac{1}{Z_m} \sum_{i=1}^{n} \mathbb{1}_{F_m(x_i)=y_i} w_{i,m+1} = \frac{1}{Z_m} \sum_{i=1}^{n} \mathbb{1}_{F_m(x_i)\neq y_i} w_{i,m+1} \quad .$$

Wieso ist dies kein Widerspruch zu dem Grundprinzip des Boosting, dass jedes Basismodell F_m die zuvor falsch klassifizierten Daten zusätzlich gewichtet?