Travaux dirigés de Machine learning

Cycle pluridisciplinaire d'études supérieures

Université Paris sciences et lettres

Joon Kwon

vendredi 5 juin 2020

Exercice 1 (*Adaboost*). — On se place dans un cadre de classification binaire, c'est-à-dire avec un ensemble d'entrées $\mathscr{X}$ quelconque, et un ensemble de sorties $\mathscr{Y} = \{-1, 1\}$. Soit $m, n \geqslant 1$ des entiers, et $\gamma \in]0, 1/2[$. Soit $S = (x_i, y_i)_{i \in [n]}$ un échantillon et $\hat{f}^{(1)}, \dots, \hat{f}^{(m)} \in \mathscr{F}(\mathscr{X}, \mathscr{Y})$ des prédicteurs. On pose $\pi^{(1)} = (1/n, \dots, 1/n)$ et pour $k \in [m]$, on définit par récurrence :

$$\begin{aligned}
\varepsilon^{(k)} &= \sum_{i=1}^{n} \pi_i^{(k)} \mathbb{1}_{\left\{y_i \neq \hat{f}^{(k)}(x_i)\right\}} \\
w^{(k)} &= \frac{1}{2} \log\left(\frac{1}{\varepsilon^{(k)}} - 1\right) \\
\pi^{(k+1)} &= \left(\frac{\pi_i^{(k)} \exp\left(-w^{(k)} y_i \hat{f}^{(k)}(x_i)\right)}{\sum_{j=1}^{n} \pi_j^{(k)} \exp\left(-w^{(k)} y_j \hat{f}^{(k)}(x_j)\right)}\right)_{i \in [n]},
\end{aligned}$$

et on suppose que $\varepsilon^{(k)} \leqslant \frac{1}{2} - \gamma$ pour tout $k \in [m]$. Pour $0 \leqslant k \leqslant m$, on pose

$$\hat{f}^{(1:k)} = \sum_{l=1}^{k} w^{(l)} \hat{f}^{(l)},$$

et considère le prédicteur $\hat{f}$ défini par :

$$\forall x \in \mathscr{X}, \quad \hat{f}(x) = \operatorname{sign}\left(\hat{f}^{(1:m)}(x)\right),$$

et on note $\varepsilon_S(\hat{f})$ son erreur d'apprentissage :

$$\varepsilon_S(\hat{f}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\left\{\hat{f}(x_i) \neq y_i\right\}}.$$

Le but de l'exercice est de montrer que $\varepsilon_S(\hat{f}) \leqslant e^{-2\gamma^2 m}$.

Pour $0 \leqslant k \leqslant m$, on définit la quantité :

$$Z^{(k)} = \frac{1}{n}\sum_{i=1}^{n} \exp\left(-y_i \hat{f}^{(1:k)}(x_i)\right).$$

1) Montrer que

$$\varepsilon_S(\hat{f}) \leqslant Z^{(m)}.$$

2) Montrer que pour $0 \leqslant k \leqslant m$,

$$\pi^{(k+1)} = \left(\frac{\exp\left(-y_i \hat{f}^{(1:k)}(x_i)\right)}{\sum_{j=1}^{n} \exp\left(-y_j \hat{f}^{(1:k)}(x_j)\right)}\right)_{i \in [n]}.$$

3) Montrer que pour $0 \leqslant k \leqslant m-1$,

$$\frac{Z^{(k+1)}}{Z^{(k)}} = 2\sqrt{\varepsilon^{(k+1)}(1-\varepsilon^{(k+1)})}.$$

4) En déduire que

$$\varepsilon_S(\hat{f}) \leqslant e^{-2\gamma^2 m}.$$