STA 212 **(Le 07/05/2022 de 08h45 à 10h45)**

Examen

Enseignant: Mohammed Sedki *pageweb:* masedki.github.io

Instructions : Tous les documents de cours, notes ou supports personnels sont autorisés. Il est interdit de faire appel à une connexion internet durant l'épreuve.

Problème : Autour de l'algorithme AdaBoost (10 points)

Soit $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ un jeu de données d'apprentissage où $x \in \mathbb{R}^d$ et $y \in \{-1, +1\}$.

Rappelons l'algorithme de classification Adaboost

I. Initialisation des poids des données $w_i^1 = \frac{1}{n}$ pour $i = 1, \ldots, n$.

II. Pour $m = 1, \ldots, M$

(a) Ajustement d'une règle faible de classification $\widehat{g}_m(x)$ sur le jeu de données d'apprentissage $(x_1, y_1), \ldots, (x_n, y_n)$ pondéré par les poids $w_1^m, \ldots, w_n^m$.

(b) Calcul de l'erreur

$$\varepsilon_m = \sum_{i=1}^{n} w_i^m \mathbb{I}\Big\{y_i \neq \widehat{g}_m(x_i)\Big\}.$$

(c) Calcul du coefficient de confiance de la règle $\widehat{g}_m(x)$

$$\alpha_m = \frac{1}{2} \ln\left(\frac{1 - \varepsilon_m}{\varepsilon_m}\right).$$

(d) Mise à jour des poids du jeu de données d'apprentissage

$$w_i^{m+1} = \frac{w_i^m \exp\big[-\alpha_m y_i \widehat{g}_m(x_i)\big]}{Z_m},$$

où $Z_m = \sum_{i=1}^n w_i^m \exp\big[-\alpha_m y_i \widehat{g}_m(x_i)\big]$ est la constante de normalisation des poids.

(e) La règle de classification ultime est donnée par $h(x) = \text{sign}\big[\widehat{g}(x)\big]$ où

$$\widehat{g}(x) = \sum_{m=1}^{M} \alpha_m \widehat{g}_m(x).$$

(1) Vérifier que

$$\alpha_m = \underset{\alpha}{\text{argmin}} \left\{ \sum_{i=1}^{n} w_i^m \exp\Big[-\alpha y_i \widehat{g}_m(x_i)\Big] \right\}.$$

(2) Vérifier l'égalité

$$Z_m = 2\sqrt{\varepsilon_m(1 - \varepsilon_m)}.$$

(3) Supposons que $\gamma \leq \frac{1}{2} - \varepsilon_m$ pour tout $m = 1, \ldots, M$. Montrer que

$$\widehat{L}_n(h) \leq \exp\big(-2\gamma^2 M\big),$$

où $\widehat{L}_n(h) = \frac{1}{n}\sum_{i=1}^n \mathbb{I}\Big\{y_i \neq h(x_i)\Big\} = \frac{1}{n}\sum_{i=1}^n \mathbb{I}\Big\{y_i \widehat{g}(x_i) \leq 0\Big\}$.[1]

1. Indication : écrire w_i^{M+1} en fonction des $\{Z_m\}_{m=1}^M$ et des $\big(y_i, \widehat{g}(x_i)\big)_{i=1}^n$.