Apprentissage statistique: TD 7
Applications of Stone's Theorem

Emile Contal
http://econtal.perso.math.cnrs.fr/teaching

7 mars 2016

Rappels.
Soit $Z_m = \big(U_1, \dots, U_m\big)$ une suite de variables aléatoires iid et $x, U \mapsto g_n(x, U)$ un classifieur *randomisé*. On rappelle la définition du *classifieur par votes* :

$$\bar{g}_n(x, Z_m) = \mathbb{1}_{\frac{1}{m}\sum_{j=1}^m g_n(x,U_j) \geq \frac{1}{2}} \,.$$

On définit également le *classifieur par espérance* :

$$\widetilde{g}_n(x) = \mathbb{1}_{\mathbf{E}_U[g_n(x,U)] \geq \frac{1}{2}} = \lim_{m \to \infty} \bar{g}_n(x, Z_m) \,.$$

Théorème 1 (Stone, 1977)**.** *Soient $W_{ni}(x)$ des poids positifs dépendants de $X_1, \dots, X_n$, tels que $\sum_{i=1}^n W_{ni}(x) = 1$. Soit (g_n) une suite de classifieurs pondérés :*

$$g_n = 1 \textit{ ssi } \sum_{i=1}^n \mathbb{1}_{Y_i=1} W_{ni}(x) \geq \sum_{i=1}^n \mathbb{1}_{Y_i=0} W_{ni}(x) \,.$$

On a que (g_n) est universellement consistante lorsque :

(1) Il existe une constante c telle que pour toute fonction f mesurable positive d'espérance finie, $\mathbf{E}\big[\sum_{i=1}^n W_{ni}(X) f(X_i)\big] \leq c\mathbf{E}[f(X)]$,

(2) Pour tout $a > 0$, $\mathbf{E}\big[\sum_{i=1}^n W_{ni}(X) \mathbb{1}_{\|X_i - X\| > a}\big] \to 0$,

(3) $\mathbf{E}\big[\max_{1 \leq i \leq n} W_{ni}(X)\big] \to 0$.

Lemme 1 (Stone, 1977)**.** *Pour toute fonction mesurable f, tout entiers n et $k \leq n$, il existe une constante $c_d > 0$ telle que :*

$$\sum_{i=1}^k \mathbf{E}\Big[\big|f\big(X_{(i)}(X)\big)\big|\Big] \leq c_d k \mathbf{E}\big[\,|f(X)|\,\big] \,,$$

avec $c_d \leq \big(1 + 2/\sqrt{2 - \sqrt{3}}\big)^d - 1$.

Exercice 1. (Consistency by Randomization)

On se place pour simplifier dans le cas $d = 1$ et on considère que le support de X n'est pas atomique. Pour un x fixé, soient $X_{(1)}(x), \ldots, X_{(n)}(x)$ les points $X_1, \ldots, X_n$ triés selon leur distance à x, et $Y_{(1)}(x), \ldots, Y_{(n)}(x)$ les données correspondantes. Soient $U = \big(U_1, \ldots, U_n\big)$ des variables aléatoires uniformes sur $[0, 1]$. On définit une règle du plus proche voisin randomisée comme suit :

$$g_n(x, U) = Y_{(i_U)}(x) \text{ où } i_U = \operatorname*{argmin}_i \max(i, mU_i)\,,$$

où $m \leq n$ est un paramètre.

Soit $\widetilde{g}_n(x) = \mathbb{1}_{\mathbf{E}_U g_n(x,U) \geq 1/2}$ le classifieur par espérance, Montrer que $\widetilde{g}_n$ est consistant lorsque $m \to \infty$ et $\frac{m}{n} \to 0$. Vous pourrez utiliser le fait que pour X indépendant des X_i, $\big\|X_{(k)}(X) - X\big\| \to 0$ avec probabilité 1 lorsque $\frac{k}{n} \to 0$.

Exercice 2. (Random Forests)

On suppose que le support de μ est $\mathcal{X} = [0, 1]^d$. Un *arbre de décision randomisé* $g_n(x, U)$ est un classifieur qui construit une partition hiérarchique de $\mathcal{X}$ sous forme d'arbre binaire dont les noeuds correspondent à une partie rectangulaire de $\mathcal{X}$. La racine de l'arbre est associée à $\mathcal{X}$ en entier. A chaque noeud, on découpe le rectangle correspondant en deux sous-rectangles suivant une certaine coordonnée et un certain seuil. L'ensemble des cellules associées aux feuilles maintient en permanence une partition de $\mathcal{X}$. Un *arbre de décision purement randomisé* de taille k est construit en répétant k fois la procédure :

- choisir une feuille uniformément parmi les feuilles de l'arbre,
- choisir une coordonnée uniformément parmi les d variables,
- choisir un seuil uniformément dans la longueur du rectangle correspondante,
- découper la feuille suivant la variable et le seuil en deux sous feuilles.

Une *forêt aléatoire* est le classifieur par votes associé à une séquence d'arbres de décision $\big(g_n(x, U)\big)_n$.

Montrer qu'une forêt purement aléatoire est consistante lorsque $k \to \infty$ et $\frac{k}{n} \to 0$.