

Apprentissage statistique: TD5

Random Forests and Averaging Classifiers

Emile Contal

<http://econtal.perso.math.cnrs.fr/teaching>

13 février 2015

Notations et définitions.

Soit $(X, Y), (X_1, Y_1), \dots, (X_n, Y_n)$ des paires iid de variables aléatoires sur $\mathbb{R}^d \times \{0, 1\}$ telles que X soit de mesure μ et $\Pr[Y = 1 | X = x] = \eta(x)$. On note $D_n = \{(X_i, Y_i)\}_{i \leq n}$ l'ensemble d'apprentissage et g_n un classifieur (où la dépendance en D_n est implicite). On notera $L(g_n) = \Pr[g_n(X) \neq Y | D_n]$ l'erreur de g_n .

Une suite de classifieur $(g_n)_n$ est *consistante* pour une distribution de (X, Y) lorsque $\mathbf{E}[L(g_n)] \rightarrow L^*$ où L^* est l'erreur de Bayes. Une suite de classifieur est *universellement consistante* si elle est consistante pour toute distribution de (X, Y) .

Soit $Z_m = (U_1, \dots, U_m)$ une suite de variables aléatoires iid. Soit $x, U \mapsto g_n(x, U)$ un classifieur *randomisé*, qui dépend également de U . On définit $g_n^{(m)}(x, Z_m) = \mathbf{1}_{\frac{1}{m} \sum_{j=1}^m g_n(x, U_j) \geq \frac{1}{2}}$ le *classifieur par votes*, et $\bar{g}_n(x) = \mathbf{1}_{\mathbf{E}_U[g_n(x, U)] \geq \frac{1}{2}} = \lim_{m \rightarrow \infty} g_n^{(m)}(x, Z_m)$ le *classifieur par espérance*.

Exercice 1. (Voting and Averaged Classifiers)

Soit $(g_n)_n$ une suite consistante de classifieurs randomisés. Montrer que son classifieur par votes ainsi que son classifieur par espérance sont également consistants.

Théorème 1 (Devroye, 1996). Soient $A_n = \{A^1, A^2, \dots\}$ une partition de \mathbb{R}^d et $A_n(x)$ la cellule de A_n contenant x . Soit (g_n) une suite de classifieurs votants sur les cellules de la partition :

$$g_n = 1 \text{ ssi } \sum_{i=1}^n \mathbf{1}_{Y_i=1} \mathbf{1}_{X_i \in A_n(x)} \geq \sum_{i=1}^n \mathbf{1}_{Y_i=0} \mathbf{1}_{X_i \in A_n(x)}.$$

On a que (g_n) est consistante lorsque :

- (1) $\sup_{x, y \in A_n(x)} \|x - y\| \rightarrow 0$ en probabilité,
- (2) $\sum_{i=1}^n \mathbf{1}_{X_i \in A_n(x)} \rightarrow \infty$ en probabilité.

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 $(g_n(x, U))_n$.

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

Théorème 2 (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 :

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

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, $\mathbf{E}[\sum_{i=1}^n W_{ni}(X) f(X_i)] \leq c \mathbf{E}[f(X)]$,
- (2) Pour tout $a > 0$, $\mathbf{E}[\sum_{i=1}^n W_{ni}(X) \mathbb{1}_{\|X_i - X\| > a}] \rightarrow 0$,
- (3) $\mathbf{E}[\max_{1 \leq i \leq n} W_{ni}(X)] \rightarrow 0$.

Exercice 3. (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 $X_{(1)}(x), \dots, X_{(n)}(x)$ les points X_1, \dots, X_n triés selon leur distance à x . Soient $U = (U_1, \dots, U_n)$ des variables aléatoires uniformes sur $[0, 1]$. On définit une règle du plus proche voisin randomisée comme suit :

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

où $m \leq n$ est un paramètre.

Montrer que le classifieur par espérance du plus proche voisin randomisé est consistant lorsque $m \rightarrow \infty$ et $\frac{m}{n} \rightarrow 0$.