

TRAVAUX DIRIGÉS N° 2 : Concentration, théorie de VC

Stephan CLÉMENÇON <stephan.clemencon@telecom-paris.fr>

Ekhine IRUOZKI <irurozki@telecom-paris.fr>

EXERCICE 1. On se place dans le cadre de la classification binaire : soient un descripteur aléatoire X à valeurs dans un espace mesurable $\mathcal{X} \subset \mathbb{R}^d$ ($d \in \mathbb{N}^*$) et un label aléatoire Y valant -1 ou 1 . On considère une classe finie \mathcal{G} de classifieurs $\mathcal{X} \rightarrow \{-1, 1\}$ telle que les deux labels sont parfaitement séparables par un élément de \mathcal{G} , i.e. $\min_{g \in \mathcal{G}} L(g) = 0$ pour le risque $L : g \in \mathcal{G} \mapsto \mathbb{P}(g(X) \neq Y) \in [0, 1]$.

Soit $n \in \mathbb{N}^*$. On suppose que l'on dispose d'un échantillon i.i.d. $\{(X_i, Y_i)\}_{1 \leq i \leq n}$ suivant la même loi que (X, Y) et on note \hat{g}_n un minimiseur de l'erreur empirique de classification :

$$\hat{g}_n \in \min_{g \in \mathcal{G}} L_n(g) \quad \text{où} \quad L_n : g \in \mathcal{G} \mapsto \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{g(X_i) \neq Y_i\}}.$$

- 1) Montrer que $\min_{g \in \mathcal{G}} L_n(g) = 0$ presque-sûrement.
- 2) Montrer que $\mathbb{P}(L(\hat{g}_n) > \epsilon) \leq |\mathcal{G}|(1 - \epsilon)^n$ pour tout $\epsilon \in [0, 1]$.

En déduire que $\mathbb{P}(L(\hat{g}_n) > \epsilon) \leq |\mathcal{G}|e^{-n\epsilon}$ pour tout $\epsilon > 0$.

Indication. Utiliser $\mathcal{G}_B := \{g \in \mathcal{G} : L(g) > \epsilon\}$ ainsi qu'une borne d'union.

- 3) Déduire de la question précédente que $\mathbb{E}(L(\hat{g}_n)) \leq \frac{\log(e|\mathcal{G}|)}{n}$.

Indication. Pour toute variable aléatoire Z positive, $\mathbb{E}(Z) = \int_0^{+\infty} \mathbb{P}(Z > t) dt$.

EXERCICE 2. On se place dans le cadre de la classification binaire. On utilisera les mêmes notations que dans l'exercice précédent. On pose $L^* := L(g^*)$ avec $g^* : x \in \mathcal{X} \mapsto 2\mathbb{1}_{\{\eta(x) \geq 1/2\}} - 1$ et on note $\eta : x \in \mathcal{X} \mapsto \mathbb{P}(Y = 1 | X = x) \in [0, 1]$ la fonction de régression. Soit $(\eta_n)_{n \in \mathbb{N}^*}$ une suite de fonctions définies sur \mathcal{X} à valeurs dans $]0, 1[$. Pour tout $n \in \mathbb{N}^*$ on considère le classifieur $g_n : x \in \mathcal{X} \mapsto 2\mathbb{1}_{\{\eta_n(x) \geq 1/2\}} - 1$.

- 1) On suppose qu'il existe $\delta > 0$ tel que $|\eta(x) - 1/2| \geq \delta$ pour tout $x \in \mathcal{X}$. Montrer que

$$L(g_n) - L^* \leq \frac{2 \mathbb{E}((\eta_n(X) - \eta(X))^2)}{\delta}.$$

2) Montrer que si $L^* = 0$, alors quel que soit $q \in [1, +\infty[$

$$L(g_n) \leq 2^q \mathbb{E} (|\eta_n(X) - \eta(X)|^q).$$

Soient maintenant $\eta' : \mathcal{X} \rightarrow]0, 1[$ et $g : x \in \mathcal{X} \mapsto 2\mathbb{1}_{\{\eta'(x) \geq 1/2\}} - 1$.

3) On suppose que $\mathbb{P}\{\eta'(X) = 1/2\} = 0$ et que $\mathbb{E} (|\eta_n(X) - \eta'(X)|) \rightarrow 0$ lorsque $n \rightarrow +\infty$. Montrer que $L(g_n) \rightarrow L(g)$ lorsque $n \rightarrow +\infty$.

4) On suppose que le label Y n'est plus observable, mais qu'une variable Z à valeurs dans $\{-1, +1\}$ l'est, telle que :

$$\mathbb{P}(Z = 1 \mid Y = -1, X) = \mathbb{P}(Z = 1 \mid Y = -1) = a < 1/2,$$

$$\mathbb{P}(Z = -1 \mid Y = 1, X) = \mathbb{P}(Z = -1 \mid Y = 1) = b < 1/2.$$

On pose à présent $\eta' : x \in \mathcal{X} \mapsto \mathbb{P}(Z = +1 \mid X = x)$. Montrer que :

$$L(g) \leq L^* \left(1 + \frac{2|a-b|}{1-2\max(a,b)} \right).$$

Que peut-on en déduire lorsque $a = b$?

EXERCICE 3. Calculer la VC dimension des classes \mathcal{A} d'ensembles suivantes :

1) $\mathcal{A} = \{]-\infty, x_1] \times \dots \times]-\infty, x_d] : (x_1, \dots, x_d) \in \mathbb{R}^d \},$

2) \mathcal{A} est constituée des rectangles de \mathbb{R}^d .

EXERCICE 4. Donner une borne supérieure de la VC dimension de la classe des boules fermées dans \mathbb{R}^d :

$$\mathcal{A} = \left\{ \left\{ x = (x_1, \dots, x_d) \in \mathbb{R}^d : \sum_{i=1}^d |x_i - a_i|^2 \leq b \right\} : a_1, \dots, a_d, b \in \mathbb{R} \right\}.$$

EXERCICE 5. Soit \mathcal{A} une classe d'ensembles de \mathbb{R}^d de VC dimension $V < +\infty$ et de coefficients d'éclatement $s(\mathcal{A}, n), \forall n \in \mathbb{N}^*$.

1) Montrer que : $\forall n \geq 1, s(\mathcal{A}, n) \leq (n+1)^V.$

2) Montrer que : $\forall n \geq V, s(\mathcal{A}, n) \leq (ne/V)^V.$

Indication. On utilisera le lemme de Sauer : $\forall n \geq 1, s(\mathcal{A}, n) \leq \sum_{k=0}^V \binom{n}{k}$.
