MAP553 Apprentissage Statistique

PC8 : VC-dimension, inégalités exponentielles

À une famille $\mathcal{H}$ de classifieurs $h : \mathcal{X} \to \{0,1\}$ on associe le coefficient d'éclatement

$$\mathbb{S}_{\mathcal{H}}(n) = \max_{x_1,\ldots,x_n \in \mathcal{X}} \mathsf{Card}\big\{(h(x_1),\ldots,h(x_n)) : h \in \mathcal{H}\big\} \leq 2^n, \quad \text{pour } n \in \mathbb{N}.$$

La VC-dimension de $\mathcal{H}$ est alors $V_{\mathcal{H}} = \sup\{n \in \mathbb{N} \colon \mathbb{S}_{\mathcal{H}}(n) = 2^n\}$.

« *$V_{\mathcal{H}}$ est le plus grand nombre de points que $\mathcal{H}$ peut éclater* »

1 VC-dimension des classifieurs linéaires dans $\mathcal{X} = \mathbb{R}^d$

Pour tout $w \in \mathbb{R}^d$, notons $h_w : \mathbb{R}^d \to \{0,1\}$, défini par

$$h_w(x) = \mathbf{1}_{\{\langle w,x\rangle > 0\}}, \quad \text{pour } x \in \mathbb{R}^d.$$

Nous allons déterminer la VC-dimension de $\mathcal{H} = \{h_w : w \in \mathbb{R}^d\}$.

1. Soit $e_1,\ldots,e_d$ la base canonique de $\mathbb{R}^d$. Montrer que pour tout $\delta \in \{0,1\}^d$ il existe $w_\delta \in \mathbb{R}^d$ tel que $h_{w_\delta}(e_i) = \delta_i$ pour $i = 1,\ldots,d$. En déduire une minoration de $V_{\mathcal{H}}$.
2. Soit $x_1,\ldots,x_{d+1} \in \mathbb{R}^d$ et $\lambda \in \mathbb{R}^{d+1}$ non nul tel que $\sum_{i=1}^{d+1} \lambda_i x_i = 0$.
   Posons $\delta = [\mathbf{1}_{\lambda_i > 0}]_{i=1,\ldots,d+1}$. En inspectant la somme $\sum_{i=1}^{d+1} \lambda_i \langle w_\delta, x_i\rangle$, montrez qu'il n'existe pas de $w_\delta \in \mathbb{R}^d$ vérifiant
   - $h_{w_\delta}(x_i) = \delta_i$, pour $i = 1,\ldots,d+1$,
   - $\langle w_\delta, x_i\rangle \neq 0$, pour $i = 1,\ldots,d+1$.
3. En déduire une majoration de $V_{\mathcal{H}}$. Conclusion ?
4. Quelle est la VC-dimension de $\tilde{\mathcal{H}} = \{h_{w,b} : w \in \mathbb{R}^d,\ b \in \mathbb{R}\}$, où $h_{w,b}(x) = \mathbf{1}_{\{\langle w,x\rangle > b\}}$?

2 Classifieurs linéaires sous contraintes de marge

Soit $\mathcal{X}$ un espace de Hilbert et $\{x_1,\ldots,x_M\} \subset \mathcal{X}$ avec $|x_i|_{\mathcal{X}} \leq A$ pour tout $i = 1,\ldots,M$. Pour $r, R > 0$ on note

$$\mathcal{H} = \big\{h_w : w \in \mathcal{X},\ |w|_{\mathcal{X}} \leq R \text{ et } |\langle w, x_i\rangle_{\mathcal{X}}| \geq r \text{ pour } i = 1,\ldots,M\big\},$$

où $h_w : \{x_1,\ldots,x_M\} \to \{0,1\}$ est défini par $h_w(x) = \mathbf{1}_{\{\langle w,x\rangle_{\mathcal{X}} > 0\}}$ pour tout $x \in \{x_1,\ldots,x_M\}$. Nous allons montrer que $V_{\mathcal{H}} \leq A^2R^2/r^2$ lorsque $M > A^2R^2/r^2$. On supposera dorénavant que $M > A^2R^2/r^2$.

1. Soit $n \leq M$ et $\varepsilon_1,\ldots,\varepsilon_n$ i.i.d. de loi uniforme sur $\{-1,1\}$. Montrer que

$$\mathbb{E}\left[|\sum_{i=1}^n \varepsilon_i x_i|_{\mathcal{X}}^2\right] = \sum_{i=1}^n \mathbb{E}\left[|\varepsilon_i x_i|_{\mathcal{X}}^2\right] \leq nA^2.$$

2. En déduire qu'il existe $y \in \{-1,1\}^n$ tel que $\left|\sum_{i=1}^n y_i x_i\right|_{\mathcal{X}}^2 \leq nA^2$.
3. Supposons qu'il existe $w \in \mathcal{X}$ tel que $h_w \in \mathcal{H}$ et $y_i\langle w, x_i\rangle_{\mathcal{X}} \geq r$ pour $i = 1, \dots, n$.

$$\text{Montrer que} \qquad nr \leq \langle w, \sum_{i=1}^n y_i x_i\rangle_{\mathcal{X}} \leq RA\sqrt{n}.$$

4. En déduire que $\mathbb{S}_{\mathcal{H}}(n) < 2^n$ pour $R^2A^2/r^2 < n \leq M$. Conclure.

3 Inégalité de Hoeffding

Soit Z une variable aléatoire à valeurs dans $[a,b]$ telle que $\mathbb{E}[Z] = 0$.

1. Montrer que pour tout $s, t > 0$ on a $\mathbb{P}(Z \geq t) \leq e^{-st}\mathbb{E}[e^{sZ}]$.
2. À l'aide des sous-questions suivantes, établir la majoration $\mathbb{E}[e^{sZ}] \leq e^{s^2(b-a)^2/8}$.
   (a) Justifier l'inégalité $e^{sz} \leq \frac{z-a}{b-a}e^{sb} + \frac{b-z}{b-a}e^{sa}$ pour tout $z \in [a,b]$.
   (b) Montrer que $\log(\mathbb{E}[e^{sZ}]) \leq g(s(b-a))$ avec $g(u) = -pu + \log(1-p+pe^u)$ où $p = -a/(b-a)$.
   (c) Montrer que $g(u) \leq u^2/8$ pour tout $u > 0$.
3. En déduire que $\mathbb{P}(Z \geq t) \leq e^{-2t^2/(b-a)^2}$.
4. Soit $X_1, \dots, X_n$ variables aléatoires réelles, indépendantes, telles que $X_i \in [a_i, b_i]$ presque sûrement. En notant $\bar{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (b_i - a_i)^2$, montrer que pour $t > 0$

$$\mathbb{P}\left(\sum_{i=1}^n (X_i - \mathbb{E}[X_i]) \geq t\right) \leq e^{-2t^2/n\bar{\sigma}^2} \quad \text{et} \quad \mathbb{P}\left(\sum_{i=1}^n (\mathbb{E}[X_i] - X_i) \geq t\right) \leq e^{-2t^2/n\bar{\sigma}^2}.$$

Apprentissage statistique en Master2

– **MVA :** Modélisation de la vision et analyse d'images.
  Etablissements : Paris 5, ENS Cachan, Polytechnique, Dauphine, Telecom, Centrale, Ponts
  http://www.math.ens-cachan.fr/version-francaise/formations/master-mva/

– **M2 Probabilités et Statistiques :** Probabilités, Statistiques et Statistiques appliquées.
  Etablissements : Orsay, Polytechnique, ENS Ulm, Agro
  http://webens-ng.math.u-psud.fr/M2/PS/

– **MathSV :** Mathématiques pour les Sciences du Vivant.
  (Ecologie, Biologie des systèmes et apprentissage statistique, Biomécanique, Neuroimagerie)
  Etablissements : Orsay, Polytechnique, ENS Cachan
  http://webens-ng.math.u-psud.fr/M2/MathSV/