Ecole Polytechnique **Statistical Learning**
Master M2 Data Science : 2026/2027 **Stephan Clémençon - Quentin Bouniot**

Travaux dirigés n° 2 : Empirical Risk, Concentration, VC Theory

---

Stephan Clémençon <stephan.clemencon@telecom-paristech.fr>
Quentin Bouniot <quentin.bouniot@telecom-paris.fr>

Exercice 1. We work in the framework of binary classification : let X be a random descriptor taking values in a measurable space $\mathcal{X} \subset \mathbb{R}^d$ ($d \in \mathbb{N}^*$), and let Y be a random label taking the value -1 or 1. Consider a finite class $\mathcal{G}$ of classifiers $\mathcal{X} \to \{-1, 1\}$ such that the two labels are perfectly separable by an element of $\mathcal{G}$, *i.e.* $\min_{g\in\mathcal{G}} \mathrm{L}(g) = 0$, for the risk $\mathrm{L} : g \in \mathcal{G} \mapsto \mathbb{P}\left(g(\mathrm{X}) \neq \mathrm{Y}\right) \in [0, 1]$.

Let $n \in \mathbb{N}^*$. Suppose that we have an i.i.d. sample $\{(\mathrm{X}_i, \mathrm{Y}_i)\}_{1\leq i\leq n}$ following the same distribution as (X, Y), and let $\hat{g}_n$ be a minimizer of the empirical classification error :

$$\hat{g}_n \in \min_{g\in\mathcal{G}} \mathrm{L}_n(g) \quad \text{where} \quad \mathrm{L}_n : g \in \mathcal{G} \mapsto \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{g(\mathrm{X}_i)\neq \mathrm{Y}_i\}}.$$

1) Show that $\min_{g\in\mathcal{G}} \mathrm{L}_n(g) = 0$ almost surely.
2) Show that $\mathbb{P}\left(\mathrm{L}(\hat{g}_n) > \epsilon\right) \leq |\mathcal{G}|(1-\epsilon)^n$ for every $\epsilon \in [0, 1]$.
   Deduce that $\mathbb{P}\left(\mathrm{L}(\hat{g}_n) > \epsilon\right) \leq |\mathcal{G}|e^{-n\epsilon}$.

   Hint. Use $\mathcal{G}_\epsilon := \{g \in \mathcal{G} : \mathrm{L}(g) > \epsilon\}$ together with a union bound.
3) Deduce from the previous question that $\mathbb{E}\left(\mathrm{L}(\hat{g}_n)\right) \leq \dfrac{\log(e|\mathcal{G}|)}{n}$.

   Hint. For every positive random variable Z, $\mathbb{E}\left(\mathrm{Z}\right) = \int_0^{+\infty} \mathbb{P}(\mathrm{Z} > t)\, dt$.

Exercice 2. We work in the framework of binary classification. We use the same notation as in the previous exercise. Set $\mathrm{L}^* := \mathrm{L}(g^*)$, where $g^* : x \in \mathcal{X} \mapsto 2\mathbb{1}_{\{\eta(x)\geq 1/2\}} - 1$, and let $\eta : x \in \mathcal{X} \mapsto \mathbb{P}(\mathrm{Y} = 1 \mid \mathrm{X} = x) \in [0, 1]$ be the regression function. Let $(\eta_n)_{n\in\mathbb{N}^*}$ be a sequence of functions defined on $\mathcal{X}$ with values in $]0, 1[$. For every $n \in \mathbb{N}^*$, consider the classifier $g_n : x \in \mathcal{X} \mapsto 2\mathbb{1}_{\{\eta_n(x)\geq 1/2\}} - 1$.

1) Suppose that there exists $\delta > 0$ such that $|\eta(x) - 1/2| \geq \delta$ for every $x \in \mathcal{X}$. Show that

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

2) Show that if $\mathrm{L}^* = 0$, then for every $q \in [1, +\infty[$,

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

Now let $\eta' : \mathcal{X} \to]0, 1[$ and $g' : x \in \mathcal{X} \mapsto 2\mathbb{1}_{\{\eta'(x)\geq 1/2\}} - 1$.

3) Suppose that $\mathbb{P}\{\eta'(\mathrm{X}) = 1/2\} = 0$ and that $\mathbb{E}\left(|\eta_n(\mathrm{X}) - \eta'(\mathrm{X})|\right) \to 0$ as $n \to +\infty$. Show that $\mathrm{L}(g_n) \to \mathrm{L}(g')$ as $n \to +\infty$.

4) Suppose that the label Y is no longer observable, but that a variable Z taking values in $\{-1,+1\}$ is observable, such that :

$$\begin{aligned}\mathbb{P}\left(\mathrm{Z}=1 \mid \mathrm{Y}=-1,\ \mathrm{X}\right) &= \mathbb{P}\left(\mathrm{Z}=1 \mid \mathrm{Y}=-1\right)=a<1/2,\\ \mathbb{P}\left(\mathrm{Z}=-1 \mid \mathrm{Y}=1,\ \mathrm{X}\right) &= \mathbb{P}\left(\mathrm{Z}=-1 \mid \mathrm{Y}=1\right)=b<1/2.\end{aligned}$$

Now set $\eta' : x \in \mathcal{X} \mapsto \mathbb{P}\left(\mathrm{Z}=+1 \mid \mathrm{X}=x\right)$. Show that :

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

What can be deduced when $a=b$?

Exercice 3. Compute the VC dimension of the following classes $\mathcal{A}$ of sets :

1) $\mathcal{A}=\left\{\ ]-\infty,x_1]\times\ldots\times]-\infty,x_d]\ :\ (x_1,\ldots,x_d)\in\mathbb{R}^d\ \right\}$,
2) $\mathcal{A}$ is the class of rectangles in $\mathbb{R}^d$.

Exercice 4. Give an upper bound for the VC dimension of the class of closed balls in $\mathbb{R}^d$:

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

Exercice 5. Let $\mathcal{A}$ be a class of sets in $\mathbb{R}^d$ with finite VC dimension $\mathrm{V}<+\infty$ and shattering coefficients $s(\mathcal{A},n)$, for every $n\in\mathbb{N}^*$.

1) Show that : $\forall n\geq 1$, $s(\mathcal{A},n)\leq(n+1)^{\mathrm{V}}$.
2) Show that : $\forall n\geq \mathrm{V}$, $s(\mathcal{A},n)\leq(ne/\mathrm{V})^{\mathrm{V}}$.

Hint. Use Sauer's lemma : $\forall n\geq 1$, $s(\mathcal{A},n)\leq\displaystyle\sum_{k=0}^{\mathrm{V}}\binom{n}{k}$.