Introduction to Statistical Learning

Exercise sheet n°2

DEFINITIONS

Let $\mathcal{F}$ be a class of bounded real-valued functions and $\mathcal{A}$ a class of subsets of $\mathbb{R}^d$.

— Bounded differences function - A real-valued function h of n variables over a metric space $\mathcal{Z}$ is said to be a function with bounded differences if there exist $c_1, \ldots, c_n > 0$ such that :

$$\sup_{z_1,\ldots,z_n,z_i' \in \mathcal{Z}} |h(z_1,\ldots,z_n) - h(z_1,\ldots,z_{i-1},z_i',z_{i+1},\ldots,z_n)| \leq c_i$$

— Martingale difference - If $V = (V_1, \ldots, V_n, \ldots)$ and $Z = (Z_1, \ldots, Z_n, \ldots)$ are two sequences of random variables we say that V is a martingale difference sequence wrt Z if for any $n \geq 1$ we have :

$$V_n \text{ is a function of } Z_1, \ldots, Z_n \text{ and } \mathbb{E}[V_{n+1}|Z_1, \ldots, Z_n] = 0$$

— The empirical Rademacher complexity of $\mathcal{F}$ wrt to the sample $D_n = \{Z_1, \ldots, Z_n\}$ is defined as :

$$\widehat{R}_n(\mathcal{F}) = \mathbb{E}\left(\sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i f(Z_i) \middle| D_n \right)$$

— The Rademacher complexity of $\mathcal{F}$ is defined as :

$$R_n(\mathcal{F}) = \mathbb{E}\big(\widehat{R}_n(\mathcal{F})\big)$$

— Trace $\mathrm{Tr}(\mathcal{A}, \mathbf{x}_1^n)$ of $\mathcal{A}$ over a set of point $\mathbf{x}_1^n = \{x_1, \ldots, x_n\}$ in $\mathbb{R}^d$:

$$\mathrm{Tr}(\mathcal{A}, \mathbf{x}_1^n) = \{A \cap \mathbf{x}_1^n \ : \ A \in \mathcal{A}\}$$

— Growth function $n \mapsto \gamma(\mathcal{A}, n)$ of $\mathcal{A}$

$$\gamma(\mathcal{A}, n) = \max_{\mathbf{x}_1^n} |\mathrm{Tr}(\mathcal{A}, \mathbf{x}_1^n)|$$

where $|\cdot|$ denotes the cardinality of the set.

— Vapnik-Chervonenkis dimension $V(\mathcal{A})$ or VC dimension of $\mathcal{A}$

$$V(\mathcal{A}) = \max n \in \mathbb{N} \ : \ s(\mathcal{A}, n) = 2^n$$

Exercise 1 - (Hoeffing's lemma) Consider Z a random variable such that : $\mathbb{E}(Z) = 0$ and $\mathbb{P}(Z \in [a, b]) = 1$ almost surely. Prove the following upper bound : for any $s > 0$,

$$\mathbb{E}\big(e^{sZ}\big) \leq \exp\left(\frac{s^2(b-a)^2}{8}\right)$$

Exercise 2 - (Bernstein inequality) Let $X_1, \dots, X_n$ independent random variables that are centered and uniformly bounded ($\forall i, |X_i| \leq c$ p.s.). Denote by $\sigma^2 = \mathrm{Var}(X_i)$. Show that $\forall t > 0$:

$$\mathbb{P}\{\sum_{i=1}^n X_i > t\} \leq \exp\left\{-\frac{t^2}{2n\sigma^2 + 2ct/3}\right\} .$$

Comment on the two typical behaviors of the tail probability of the empirical average that are captured by this exponential upper bound. Compare to Hoeffding's inequality.

Hint : first prove Bennett's inequality ;

$$\mathbb{P}\{\sum_{i=1}^n X_i > t\} \leq \exp\left\{-\frac{n\sigma^2}{c^2}\phi\left(\frac{ct}{n\sigma^2}\right)\right\}$$

where $\phi(u) = (1+u)\log(u) - u$.

Exercise 3 - (Azuma's inequality) Let V a martingale difference sequence wrt Z and $c_n \geq 0$ such that $Z_n \leq V_n \leq Z_n + c_n$.
Show that for any $t > 0$:

$$\mathbb{P}\left(\sum_{i=1}^n V_i > t\right) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right)$$

and

$$\mathbb{P}\left(\sum_{i=1}^n V_i < -t\right) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right)$$

Exercise 4 - (McDiarmid's inequality) Let h be a function with bounded differences with bounding constants $c_1, \dots, c_n$, and $Z_1, \dots, Z_n$ random variables.
Show that for any $t > 0$:

$$\mathbb{P}\left(h(Z_1, \dots, Z_n) - \mathbb{E}\big(h(Z_1, \dots, Z_n)\big) > t\right) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right)$$

and

$$\mathbb{P}\left(h(Z_1, \dots, Z_n) - \mathbb{E}\big(h(Z_1, \dots, Z_n)\big) < -t\right) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right)$$

Hint : introduce $V_i = \mathbb{E}[h|Z_1, \dots, Z_i] - \mathbb{E}[h|Z_1, \dots, Z_{i-1}]$ *and apply Azuma's inequality.*

Exercise 5 - (Application of McDiarmid's concentration inequality) Let $\mathcal{F}$ be a class of $[0,1]$-valued functions. Show that, with probability at least $1-\delta$:

$$R_n(\mathcal{F}) \le \widehat{R}_n(\mathcal{F}) + \sqrt{\frac{\log(1/\delta)}{2n}}$$

and also that :

$$\sup_{f\in\mathcal{F}} \left(\mathbb{E}\big(f(Z_1)\big) - \frac{1}{n}\sum_{i=1}^{n} f(Z_i) \right) \le 2\widehat{R}_n(\mathcal{F}) + 3\sqrt{\frac{\log(2/\delta)}{2n}}$$

Exercise 6 - (Massart's lemma) Let C be a finite subset of $\mathbb{R}^n$ and $R = \max_{z\in C} \|z\|_2$, where $z = (z_1, \ldots, z_n)$. Let $\varepsilon_1, \ldots, \varepsilon_n$ a sample of i.i.d. Rademacher random variables. Show that :

$$\mathbb{E}\left[\sup_{z\in C} \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i z_i\right] \le \frac{R\sqrt{2\log|C|}}{n}$$

Exercise 7 - (Sauer's lemma) Consider $\mathcal{A}$ a class of subsets of $\mathbb{R}^d$ with VC dimension $V < +\infty$ and growth function $\gamma(\mathcal{A}, n)$, $\forall n \ge 1$. Show that :

$$\forall n \ge 1\ , \quad \gamma(\mathcal{A}, n) \le \sum_{i=0}^{V} \binom{n}{i} \ .$$

Exercice 8 - (VC dimension of half-spaces) Consider the class $\mathcal{A}$ of half-spaces in $\mathbb{R}^d$ and show that its VC dimension $V(\mathcal{A}) = d+1$.

Hint : first prove the upper bound by Radon's theorem, and then build a separating hyperplane for any arbitrary labeling for some set of $d+1$ points.

Radon's theorem : Any set X of $d+2$ points in $\mathbb{R}^d$ can be partitioned into two subsets X_1 and X_2 such that the convex hulls of X_1 and X_2 intersect.

Exercise 9 - Assume that $\mathcal{T}$ is an m-dimensional vector space of real-valued functions. Compute the VC dimension of the class of sets :

$$\mathcal{A} = \{\ \{x \in \mathbb{R}^d\ :\ t(x) \ge 0\}\ :\ t \in \mathcal{T}\ \}$$

Application : what is the VC dimension of the class of closed balls in $\mathbb{R}^d$:

$$\left\{ x = (x_1, \ldots, x_d)^T \in \mathbb{R}^d\ :\ \sum_{i=1}^{d} |x_i - a_i|^2 \le b \right\}$$

where $a_1, \ldots, a_d, b \in \mathbb{R}$.

Exercice 10 - Compute the VC dimension $V(\mathcal{A})$ in the following cases :

(a) $\mathcal{A} = \{\ ]-\infty, x_1] \times \ldots \times]-\infty, x_d]\ :\ (x_1, \ldots, x_d) \in \mathbb{R}^d\ \}$,
(b) $\mathcal{A}$ is the class of all rectangles of $\mathbb{R}^2$ with axis-orthogonal edges.
(c) $\mathcal{A}$ is the class of all rectangles of $\mathbb{R}^2$.
(d) $\mathcal{A}$ is the class of all triangles of $\mathbb{R}^2$.
