Introduction to Statistical Learning

Exercise set 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 :

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

— McDiarmid's concentration inequality - Assume h is a function with bounded differences with bounding constants $c_1, \ldots, c_n$ then, we have, for any $t > 0$

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

and

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

— 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

1. (Hoeffding'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 $s > 0$,

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

2. (Hoeffing's inequality) Consider $Z_1, \ldots, Z_n$ IID over $[0,1]$ and $\overline{Z}_n = \frac{1}{n}\sum_{i=1}^{n} Z_i$. Show that we have, for any $t > 0$

$$\mathbb{P}\{\overline{Z}_n - \mathbb{E}(Z_1) > t\} \leq \exp(-2nt^2)$$

and

$$\mathbb{P}\{\overline{Z}_n - \mathbb{E}(Z_1) < -t\} \leq \exp(-2nt^2)$$

Exercise 2

1. (Azuma's inequality) Consider $V = (V_1, \ldots, V_n, ...)$ and $Z = (Z_1, \ldots, Z_n, ...)$ two sequences of random variables. We assume the following : for any $n \geq 1$,
   - V_n is a function of $Z_1, \ldots, Z_n$
   - $\mathbb{E}(V_{n+1} \mid Z_1, \ldots, Z_n) = 0$
   - there exists U_n a measurable function of $Z_1 \ldots, Z_{n-1}$ and $c_n \geq 0$ such that : $U_n \leq V_n \leq U_n + c_n$.

   Prove 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

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

2. (McDiarmid's inequality) Use the previous question to prove McDiarmid's inequality.

Exercise 3 - (Application of McDiarmid's concentration inequality) Let $\mathcal{F}$ be a class of $[0,1]$-valued functions. Show that, with probability at least $1-\delta$:

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

and also that :

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

Exercise 4 - (Sauer's lemma) Consider $\mathcal{A}$ a class of subsets of $\mathbb{R}^d$ with VC dimension $V < +\infty$ and growth function $\gamma(\mathcal{A}, n)$, $\forall n \geq 1$. Show that :

$$\forall n \geq 1 \ , \quad \gamma(\mathcal{A}, n) \leq \sum_{i=0}^{V} \left(\begin{array}{c} n \\ i \end{array} \right) \ .$$

Exercice 5 - (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.

Exercice 6 - 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 $\mathbb{R}^2$.

Exercise 7 - (Relation between Rademacher average and combinatorial complexities) Consider a class $\mathcal{G}$ of binary valued functions with shattering coefficient function denoted by $\gamma(\mathcal{G}, n)$. Show that :

$$\mathbb{E}(\widehat{R}_n(\mathcal{G})) \leq \sqrt{\frac{2\ln(\gamma(\mathcal{G}, n))}{n}}$$

and if $V(\mathcal{G}) < +\infty$ then show that, for some constant C, we have :

$$\mathbb{E}(\widehat{R}_n(\mathcal{G})) \leq C\sqrt{\frac{V(\mathcal{G})\log(n)}{n}}$$