Introduction to Statistical Learning

Mid-term exam

Duration : 2h - Lecture notes not allowed

Reminder on main definitions and results

— Union bound : $\mathbb{P}\{A \cup B\} \leq \mathbb{P}\{A\} + \mathbb{P}\{B\}$ where A and B are events.

— IID means Independent and Identically Distributed.

— Law of iterated expectation : $\mathbb{E}(U) = \mathbb{E}(\mathbb{E}(U \mid V))$ where U, V are random variables.

— Subadditivity of supremum operator : $\sup(f+g) \leq \sup(f) + \sup(g)$ and $\sup(f) - \sup(g) \leq \sup(f-g)$.

Exercise 1

1. Consider Q a real-valued random variable such that : $\mathbb{E}(Q) = 0$ and $\mathbb{P}(Q \in [a,b]) = 1$. Prove the following upper bound : for any $s > 0$,

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

2. Consider $V = (V_1, \ldots, V_n, ...)$ and $Z = (Z_1, \ldots, Z_n, ...)$ two sequences of real-valued random variables. We assume the following : for any $n \geq 1$,

   - $V_n = \psi(Z_1, \ldots, Z_n)$ for some mesurable function ψ
   - $\mathbb{E}(V_{n+1} \mid Z_1, \ldots, Z_n) = 0$
   - there exists a sequence T_n which is a measurable function of $(Z_1, \ldots, Z_{n-1})$ and $c \geq 0$ such that : for any n, $T_n \leq V_n \leq T_n + c$

   Find the expression of $\kappa(t,n,c)$ such that, for any $t > 0$

$$\mathbb{P}\left(\sum_{i=1}^n V_i > t\right) \leq \kappa(t,n,c), \text{ and } \mathbb{P}\left(\sum_{i=1}^n V_i < -t\right) \leq \kappa(t,n,c)\ .$$

3. Consider a real-valued and measurable function h of n variables such that there exist $c > 0$ such that : for any $i \in \{1, \ldots, n\}$

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

   Assume that $Z_1, \ldots, Z_n$ are IID random variables. Prove that, for any $t > 0$

$$\mathbb{P}\left(h(Z_1, \ldots, Z_n) - \mathbb{E}\big(h(Z_1, \ldots, Z_n)\big) > t\right) \leq \kappa(t,n,c)$$

   and

$$\mathbb{P}\left(h(Z_1, \ldots, Z_n) - \mathbb{E}\big(h(Z_1, \ldots, Z_n)\big) < -t\right) \leq \kappa(t,n,c)$$

   where $\kappa(t,n,c)$ is as before.

Exercise 2 - Let $\mathcal{F}$ a class of functions from $\mathbb{R}^d$ to $[-B, B]$, with $B > 0$. Consider random sign variables $\varepsilon_1, \ldots, \varepsilon_n$ IID such that $\mathbb{P}(\varepsilon_1 = -1) = \mathbb{P}(\varepsilon_1 = +1) = 1/2$. Consider the *empirical Rademacher complexity* defined as

$$\widehat{R}_n(\mathcal{F}) = \frac{1}{n}\mathbb{E}\left(\sup_{f\in\mathcal{F}} \sum_{i=1}^{n} \varepsilon_i f(X_i) \middle| X_1, \ldots, X_n\right)$$

and the *average Rademacher complexity* as :

$$\bar{R}_n(\mathcal{F}) = \frac{1}{n}\mathbb{E}\left(\sup_{f\in\mathcal{F}} \sum_{i=1}^{n} \varepsilon_i f(X_i)\right)$$

1. Show that for a given class $\mathcal{F}$ of functions, the empirical Rademacher complexity seen as a function of $X_1, \ldots, X_n$ satisfies the bounded differences condition.
2. Provide an upper bound on the average Rademacher complexity in terms of the empirical Rademacher complexity that holds with high probability.

Exercise 3 - Consider the binary classification model where the random pair (X, Y) has distribution P over $\mathbb{R}_+ \times \{0, 1\}$ and :

— the marginal distribution of X over $\mathbb{R}_+$ is denoted P_X

— the conditional distribution of Y given $X = x$ is a Bernoulli distribution with parameter $\eta(x) = \dfrac{\theta}{x+\theta}$, for any $x \in \mathbb{R}_+$, and for fixed $\theta > 0$.

1. Assume that the marginal distribution follows a uniform distribution $P_X = \mathcal{U}([0, \alpha\theta])$ over $\mathbb{R}_+$ with $\alpha > 1$.
   (a) Find the minimizing argument g^* of $L(g) = \mathbb{P}(Y \neq g(X))$ over all measurable classifiers $g : \mathbb{R}_+ \to \{0, 1\}$.
   (b) Compute $L^* = L(g^*)$ in the case where the marginal distribution $P_X = \mathcal{U}([0, \alpha\theta])$ with $\alpha > 1$.
2. Now assume we have the following IID data $(X_1, Y_1), \ldots, (X_n, Y_n)$ available.
   (a) Assuming P_X as before (with α fixed), propose an empirical estimate $\widehat{\theta}$ of θ based only on $X_1, \ldots, X_n$.
   (b) For general P_X over $\mathbb{R}_+$ (unspecified, not necessarily uniformly distributed, and not depending on θ), what is a possible empirical estimate $\widehat{\theta}$ for θ ?
   (c) Denote by $\widehat{\eta}$ the plugin estimate of η based on $\widehat{\theta}$, what is the plugin classifier $\widehat{g}$ based on $\widehat{\eta}$? Find a bound on $L(\widehat{g}) - L^*$ depending on the quantity $\mathbb{E}(|\widehat{\eta}(X) - \eta(X)|)$.

Exercice 4 - Find the optimal elements h^* and $\mathcal{L}^* = \mathcal{L}(h^*)$ in the following cases of error measures with binary classification data (request : all notations should be made explicit in your solutions) :

1. $\mathcal{L}(h) = \mathbb{P}\big(h(X) \neq Y\big)$.
2. $\mathcal{L}(h) = \mathbb{P}(h(X) \neq Y)$ with the constraint $\mathbb{P}(h(X) = 1) = u$ with $u \in (0, 1)$.
3. $\mathcal{L}(h) = \mathbb{E}\big(\exp(-Y \cdot h(X))\big)$.