Homework – 1

Exercise 1:[Rademacher complexities] Let $A \subset \mathbb{R}^n$ denote a subset (not necessarily a vector space!) of vectors $a = (a_1, \ldots, a_n)^\top$, and define the Rademacher complexity of the set A as

$$\mathcal{R}(A) = \mathbb{E}_\epsilon \left[\sup_{a \in A} \langle a, \epsilon \rangle \right],$$

where $\epsilon = (\epsilon_1, \ldots, \epsilon_n)^\top$, and the ϵ_is are independent Rademacher random variables.

1. Basic facts:
   (a) Justify that $\mathcal{R}(A) = \mathcal{R}(-A)$.
   (b) Show that $\mathcal{R}(A) \geq 0$.
   (c) Prove that $\mathcal{R}(A \cup -A) = \mathbb{E}_\epsilon \left[\sup_{a \in A} |\langle a, \epsilon \rangle| \right]$, and give an example to justify that in general $\mathcal{R}(A \cup -A) \neq \mathcal{R}(A)$.
2. Show the next properties:
   (a) $A \subset B$ implies $\mathcal{R}(A) \leq \mathcal{R}(B)$.
   (b) $\mathcal{R}(cA + \{b\}) = |c| \, \mathcal{R}(A)$.
   (c) $\mathcal{R}(A + B) = \mathcal{R}(A) + \mathcal{R}(B)$.
   (d) $\mathcal{R}(Conv(A)) = \mathcal{R}(A)$, with $Conv(A) = \{\sum_{i=1}^n \theta_i a_i \mid (\theta_i)_{1 \leq i \leq n} \in (\mathbb{R}_+)^n, \sum_{i=1}^n \theta_i = 1\}$.
   Hint: Use that with nonnegative scalars θ_is and a real-valued function f, ‘ $\sup_{u_1, u_2} \sum_{i=1}^2 \theta_i f(u_i) = \sum_{i=1}^2 \theta_i \sup_{u_i} f(u_i)$.
3. Bounded-difference inequality (BDI):

   Theorem [BDI]: *Let $f : \mathcal{X}^n \to \mathbb{R}$ a measurable function such that there exist numeric constants $c_1, \ldots, c_n \in \mathbb{R}_+$ such that, for any $1 \leq i \leq n$,*

$$\sup_{x_1, \ldots, x_n, x_i' \in \mathcal{X}} |f(x_1, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n)| \leq c_i.$$

   Then for every $t > 0$,

$$\mathbb{P}\left[f(D_n) - \mathbb{E}\left[f(D_n) \right] > t \right] \vee \mathbb{P}\left[f(D_n) - \mathbb{E}\left[f(D_n) \right] < -t \right] \leq e^{-\frac{2t^2}{\sum_{i=1}^n c_i^2}}.$$

   where $D_n = (X_1, \ldots, X_n)$ denotes a n-tuple of independent random variables.

   Let $\mathcal{F}$ be a class of real-valued functions such that $\sup_{f \in \mathcal{F}} \|f\|_\infty \leq M < +\infty$, and let us note $\phi(D_n) = \sup_{f \in \mathcal{F}} \left\{ \frac{1}{n} \sum_{i=1}^n \left(\mathbb{E}\left[f(X_i) \right] - f(X_i) \right) \right\} = \sup_{f \in \mathcal{F}} \{(P - P_n) f\}$.
   (a) Prove that $|\phi(D_n) - \phi(D_n'(i))| \leq 2M/n$ a.s., for all $1 \leq i \leq n$, where $D_n'(i) = (X_1, \ldots, X_{i-1}, X_i', X_{i+1}, \ldots, X_n)$ is equal to D_n up to the ith coordinate where X_i is replaced by a copy X_i' (same probability distribution as X_i).
   (b) Using the bounded-difference inequality (BDI), prove that for every $t > 0$,

$$\mathbb{P}\left[\sup_f (P - P_n) f - \mathbb{E}\left[\sup_f (P - P_n) f \right] > t \right] \leq e^{-\frac{2nt^2}{(2M)^2}}$$
$$\mathbb{P}\left[\sup_f (P_n - P) f - \mathbb{E}\left[\sup_f (P_n - P) f \right] > t \right] \leq e^{-\frac{2nt^2}{(2M)^2}}$$

4. Symmetrization:

   With the above notation,

(a) use convexity of the supremum and Jensen's inequality to justify that

$$\mathbb{E}\left[\sup_f (P - P_n)f\right] \leq \mathbb{E}\left[\sup_f \left\{\frac{1}{n}\sum_{i=1}^{n}(f(X_i) - f(X_i'))\right\}\right],$$

where X_i' is an independent copy of X_i.

(b) show that $f(X_i) - f(X_i') \stackrel{\mathcal{L}}{=} \epsilon_i\,[\,f(X_i) - f(X_i')\,]$, and deduce that the joint distribution of $(f(X_i) - f(X_i'))_{1\leq i\leq n}$ is equal to the one of $(\epsilon_i\,[\,f(X_i) - f(X_i')\,])_{1\leq i\leq n}$.

(c) deduce using the triangle inequality that

$$\mathbb{E}\left[\sup_f (P - P_n)f\right] \leq 2\mathbb{E}_{D,\epsilon}\left[\sup_f \left\{\frac{1}{n}\sum_{i=1}^{n}\epsilon_i f(X_i)\right\}\right].$$

(d) finally considering the set of n-dimensional vectors $\mathcal{F}(D_n) = \{(f(X_1), \ldots, f(X_n)) \in \mathbb{R}^n \mid f \in \mathcal{F}\}$, conclude that

$$\mathbb{E}\left[\sup_f (P - P_n)f\right] \leq 2\mathbb{E}_D\left[\mathcal{R}(\mathcal{F}(D_n)/n)\right].$$

REMARK: $\mathcal{R}(\mathcal{F}(D_n)/n)$ is a function of $D_n = (X_1, \ldots, X_n)$ called the *empirical* Rademacher complexity of the function class $\mathcal{F}$ at D_n. Its expectation with respect to D_n is called the Rademacher complexity of the function class $\mathcal{F}$.

□

Exercise 2:

1. ψ_2-Orlicz norm:

   Given the definition of the ψ_2-Orlicz norm discussed along the class, we aim at proving that this is truly a norm.

   (a) Using $x \mapsto e^x$ is convex on $\mathbb{R}$ combined with Jensen's inequality, prove that $\|X\|_{\psi_2} = 0$ implies

   $$E\left[X^2\right] \leq t^2\ln(2), \qquad \forall t > 0.$$

   Deduce that $\|X\|_{\psi_2} = 0$ implies $X = 0$ a.s.

   (b) For all $\lambda \in \mathbb{R}$, prove that $\|\lambda X\|_{\psi_2} = |\lambda|\,\|X\|_{\psi_2}$

   (c) Assume X and Y have finite ψ_2-norm. Using the convexity of $u \mapsto e^{u^2}$ on $\mathbb{R}$, prove that

   $$\exp\left(\frac{X+Y}{\|X\|_{\psi_2} + \|Y\|_{\psi_2}}\right)^2 \leq \frac{\|X\|_{\psi_2}}{\|X\|_{\psi_2} + \|Y\|_{\psi_2}} e^{\left(\frac{X}{\|X\|_{\psi_2}}\right)^2} + \frac{\|X\|_{\psi_2}}{\|X\|_{\psi_2} + \|Y\|_{\psi_2}} e^{\left(\frac{Y}{\|Y\|_{\psi_2}}\right)^2}$$

   Deduce that $\|X+Y\|_{\psi_2} \leq \|X\|_{\psi_2} + \|Y\|_{\psi_2}$.

   REMARK: Following the same arguments, the ψ_1-Orlicz norm can be proved to be a norm as well.

2. Centering lemma:

   By carefully checking the proofs made along the first class on sub-Gaussian random variables, it comes out that any random variable (not necessarily centered) with a finite ψ_2-norm satisfies the following inequality.

   $$\|X\|_k \leq \sqrt{\frac{e}{2}}\,\|X\|_{\psi_2}\,k, \quad \forall k \in \mathbb{N}^*.$$

   The purpose is to prove this upper bound as well as studying a few consequences.

   Assume that X is a real-valued random variable with finite ψ_2-norm.

(a) Prove that it implies, for every $t > 0$, that

$$\mathbb{P}[\,|X| > t\,] \leq 2e^{-\frac{t^2}{\|X\|_{\psi_2}}}.$$

(b) Following the same strategy as in the lecture notes, show that

$$\forall k \in \mathbb{N}^*, \qquad \|X\|_k \leq \frac{k^{1/k}}{\sqrt{2}} \|X\|_{\psi_2} \sqrt{k} \leq \frac{\sqrt{e}}{\sqrt{2}} \|X\|_{\psi_2} \sqrt{k},$$

where the last inequality comes from $\sup_{x>0} \log(x)/x \leq 1/2$.

(c) Prove that $\|\mathbb{E}[\,X\,]\|_{\psi_2} \leq \frac{\mathbb{E}[\,X\,]}{\sqrt{\log 2}}$.

(d) Deduce that there exists $c_2 > 0$ such that $\|\mathbb{E}[\,X\,]\|_{\psi_2} \leq c_2 \|X\|_{\psi_2}$.

(e) Combine the previous results to show that

$$\|X - E[\,X\,]\|_{\psi_2} \leq (1 + c_2) \|X\|_{\psi_2}.$$

Consequently, upper bounding the ψ_2-norm of a non-centered random variable allows to get a bound on its centered counterpart, up to a numeric constant.

Remark: Similarly with the ψ_1-norm, a similar result is available as well. In particular, any random variable with a finite ψ_1-norm satisfies that there exists $c_1 > 0$ such that

$$\|X\|_k \leq c_1 \|X\|_{\psi_1} k, \quad \forall k \in \mathbb{N}^*.$$

Then,

- $\|\mathbb{E}[\,X\,]\|_{\psi_1} \leq \frac{\mathbb{E}[\,X\,]}{\log 2} \leq \overbrace{\frac{c_1}{\log 2}}^{>1} \|X\|_{\psi_1}$.
- This implies that

$$\|X - E[\,X\,]\|_{\psi_1} \leq \left(1 + \frac{c_1}{\log 2}\right) \|X\|_{\psi_1}.$$

$\square$