Theory of Machine Learning

Exercise sheet 10 — Session 10

Exercise I (dual norms) ✎. Let $p \in [1, +\infty)$ and q such that $1/p + 1/q = 1$. Recall that

$$\forall u \in \mathbb{R}^d, \qquad \|u\|_p = \left(\sum_{i=1}^d |u_i|^p\right)^{1/p},$$

and that, for any norm Ω on $\mathbb{R}^d$, we define the *dual norm* of Ω by

$$\forall u \in \mathbb{R}^d, \qquad \Omega^\star(u) := \sup_{\theta : \Omega(\theta) \leq 1} u^\top \theta .$$

The goal of this exercise is to show that the dual norm of $\|\cdot\|_p$ is $\|\cdot\|_q$.

1. Show that
$$\sup_{\|\theta\|_p \leq 1} u^\top \theta \leq \|u\|_q .$$
2. Define θ coordinate-wise as $\theta_i := \operatorname{sign}(u_i) |u_i|^{q-1}$. Show that $u^\top \theta = \|u\|_q^q$.
3. With the previous θ, show that $\|\theta\|_p = \|u\|_q^{q-1}$.
4. Set $v := \theta / \|\theta\|_p$. Show that $u^\top v = \|u\|_q$ and conclude.

Exercise II (Massart's lemma) ✎. Let us assume that $\mathcal{G}$ is *finite*, that is, $\mathcal{G} = \{g_1, \ldots, g_m\}$. Let us assume further that $\frac{1}{n}\sum_{i=1}^n g_j(X_i)^2 \leq R^2$ for all $j \in [d]$. Show that the Rademacher complexity of the function class $\mathcal{G}$ satisfies

$$R_n(\mathcal{G}) \leq \sqrt{\frac{2\log m}{n}} R .$$

For simplicity's sake, we consider the X_is fixed.

1. Given $\lambda > 0$, show that
$$\exp\left(\lambda \mathbb{E}_\varepsilon\left[\sup_{g\in\mathcal{G}} \sum_{i=1}^n \varepsilon_i g(X_i)\right]\right) \leq \sum_{g\in\mathcal{G}} \mathbb{E}_\varepsilon\left[\exp\left(\lambda \sum_{i=1}^n \varepsilon_i g(X_i)\right)\right] .$$
(*Hint: Jensen's inequality and property of* sup.)
2. Show that
$$\sum_{g\in\mathcal{G}} \mathbb{E}_\varepsilon\left[\exp\left(\lambda \sum_{i=1}^n \varepsilon_i g(X_i)\right)\right] = \sum_{g\in\mathcal{G}} \prod_{i=1}^n \frac{1}{2}\left(\exp\left(\lambda g(X_i)\right) + \exp\left(-\lambda g(X_i)\right)\right).$$
(*Hint*: independence of the ε_is and direct computation of the remaining expectation.)
3. Using $\frac{e^x + e^{-x}}{2} \leq e^{x^2/2}$ (for all $x \in \mathbb{R}$), show that
$$\sum_{g\in\mathcal{G}} \prod_{i=1}^n \frac{1}{2}\left(\exp\left(\lambda g(X_i)\right) + \exp\left(-\lambda g(X_i)\right)\right) \leq \sum_{g\in\mathcal{G}} \exp\left(\frac{\lambda^2}{2}\sum_{i=1}^n g(X_i)^2\right) .$$
4. Show that
$$\sum_{g\in\mathcal{G}} \exp\left(\frac{\lambda^2}{2}\sum_{i=1}^n g(X_i)^2\right) \leq m \exp\left(\frac{n\lambda^2 R^2}{2}\right) .$$
(*Hint*: refer to the assumptions in the exercise statement on g and $\mathcal{G}$.)

5. By putting everything together, show that

$$R_n(\mathcal{G}) \leq \frac{1}{n\lambda} \log m + \frac{\lambda R^2}{2} \,.$$

6. Show that $\lambda^\star = \frac{1}{R}\sqrt{\frac{2 \log m}{n}}$ minimizes the previous bound.

7. Show the Massart's lemma bound.

Exercise III (contraction principle). Let us assume that ϕ_i, ψ_i for $i \in [n]$ are functions on Θ such that for each $i \in [n]$, $\theta, \theta' \in \Theta$ and $L > 0$,

$$|\phi_i(\theta) - \phi_i(\theta')| \leq L \, |\psi_i(\theta) - \psi_i(\theta')| \,.$$

Prove that

$$\mathbb{E}\left[\sup_{\theta \in \Theta} \left\{ \sum_{i=1}^n \varepsilon_i \phi_i(\theta) \right\}\right] \leq \mathbb{E}\left[\sup_{\theta \in \Theta} \left\{ \sum_{i=1}^n \varepsilon_i \psi_i(\theta) \right\}\right] .$$

Hint: reason by induction on n, and take expectation on ε_{n+1}.