S&DS 602: Homework 9

Due Wednesday, November 6 at 2PM, via Gradescope

1. (a) Let $\mathcal{F}$ be the class of indicators of all half-spaces on $\mathbb{R}^d$, i.e. functions

$$f(x) = \mathbf{1}\{x^\top t \geq c\} \text{ for some } t \in \mathbb{R}^d,\ c \in \mathbb{R}.$$

Show that $\mathrm{vc}(\mathcal{F}) = d + 1$.

[Hint: Show that $0, e_1, \ldots, e_d$ is shattered by $\mathcal{F}$, where $e_i \in \mathbb{R}^d$ is the i^{th} standard basis vector. Conversely, for any $x_1, \ldots, x_{d+2} \in \mathbb{R}^d$, there exist $a_1, \ldots, a_{d+2}$ not all 0 such that $a_1 x_1 + \ldots + a_{d+2} x_{d+2} = 0$ and $a_1 + \ldots + a_{d+2} = 0$. Use this to show that $x_1, \ldots, x_{d+2}$ is not shattered by $\mathcal{F}$.]

(b) Let $\mathcal{F}$ be the class of indicators of all polygons on $\mathbb{R}^2$. Show that $\mathrm{vc}(\mathcal{F}) = \infty$.

[Hint: Consider any set of points on the unit circle.]

2. Show that if $\mathcal{F}$ is any class of functions $f : \mathcal{X} \to \{0, 1\}$ for which $\mathrm{vc}(\mathcal{F}) = \infty$, then for any $c \in (0, 1/2)$ and $n \geq 1$, there exists a probability distribution P on $\mathcal{X}$ such that

$$\mathbb{E}_{X_1, \ldots, X_n \overset{iid}{\sim} P} \sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n f(X_i) - \mathbb{E} f(X_i) > c.$$

3. (Contraction principle) (a) Let $T \subset \mathbb{R}^2$ be bounded, and let $\varphi : \mathbb{R} \to \mathbb{R}$ be 1-Lipschitz. Show that

$$\sup_{t \in T}(t_1 + \varphi(t_2)) + \sup_{t \in T}(t_1 - \varphi(t_2)) \leq \sup_{t \in T}(t_1 + t_2) + \sup_{t \in T}(t_1 - t_2)$$

(b) Let $\varphi_1, \ldots, \varphi_n : \mathbb{R} \to \mathbb{R}$ be 1-Lipschitz, and let $\varepsilon_1, \ldots, \varepsilon_n$ be i.i.d. Rademacher random variables. For any bounded $T \subset \mathbb{R}^n$, show that

$$\mathbb{E} \sup_{t \in T} \sum_{i=1}^n \varepsilon_i \varphi_i(t_i) \leq \mathbb{E} \sup_{t \in T} \sum_{i=1}^n \varepsilon_i t_i.$$

[Hint: Apply (a) conditional on all but one ε_i.]

(c) Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be i.i.d., where $X_i \in \mathcal{X}$ and $Y_i \in \{0, 1\}$. Let $\mathcal{F}$ be a class of functions $f : \mathcal{X} \to \{0, 1\}$. Using Rademacher symmetrization and (b), show that

$$\mathbb{E} \sup_{f \in \mathcal{F}} \sum_{i=1}^{n} \Big(\mathbf{1}\{f(X_i) \neq Y_i\} - \mathbb{P}[f(X_i) \neq Y_i] \Big) \leq 2\, \mathbb{E} \sup_{f \in \mathcal{F}} \sum_{i=1}^{n} \varepsilon_i f(X_i).$$

4. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be i.i.d., where $X_i, Y_i \in [0, 1]$. Let

$$\mathcal{F} = \{f : [0, 1] \to [0, 1],\ f \text{ is 1-Lipschitz}\}.$$

Define

$$R(f) = \mathbb{E}[(f(X_i) - Y_i)^2], \qquad R_n(f) = \frac{1}{n} \sum_{i=1}^{n} (f(X_i) - Y_i)^2,$$

$$f_* = \arg\min_{f \in \mathcal{F}} R(f), \qquad \hat{f} = \arg\min_{f \in \mathcal{F}} R_n(f).$$

(You may assume these minimizers are unique, the latter with probability 1.) Show that for a universal constant $C > 0$, $\mathbb{E}[R(\hat{f})] \leq R(f_*) + C/\sqrt{n}$.

[Hint: Show that $X_f := R_n(f) - R(f)$ is subgaussian with respect to $d(f, g) = \frac{C\|f-g\|_\infty}{\sqrt{n}}$, and apply Dudley's inequality to bound $\mathbb{E} \sup_{f \in \mathcal{F}} |R_n(f) - R(f)|$.]