CSCI 678: Theoretical Machine Learning
Homework 1

Fall 2024, Instructor: Haipeng Luo

This homework is due on **9/22, 11:59pm**. *See course website for more instructions on finishing and submitting your homework as well as the late policy. Total points: 50*

1. (**Rademacher complexity and Dudley entropy integral**) Consider inputs $x_1, \ldots, x_n \in \mathbb{R}^d$ and the linear class $\mathcal{F} = \left\{ f_\theta(x) = \langle \theta, x \rangle \mid \theta \in \mathbb{R}^d, \|\theta\|_2 \leq b \right\}$.

   (a) (5pts) Prove the following:

   $$\widehat{\mathcal{R}}^{\text{iid}}(\mathcal{F}; x_{1:n}) \leq \frac{b}{n} \sqrt{\sum_{t=1}^{n} \|x_t\|_2^2}$$

   using the definition of Rademacher complexity directly (that is, without invoking its upper bounds in terms of covering numbers or other measures). Hint: you will need to use the inequality $\mathbb{E}[a] \leq \sqrt{\mathbb{E}[a^2]}$ for any $a \geq 0$ (which is a consequence of Jensen's inequality).

(b) (3pts) In Lecture 4, we will prove the following log covering number bound for this class: $\ln \mathcal{N}_2(\mathcal{F}|_{x_{1:n}}, \alpha) \leq \frac{b^2 \ln(2d) \sum_{t=1}^n \|x_t\|_2^2}{n\alpha^2}$. Use this bound and the Dudley entropy integral to prove

$$\widehat{\mathcal{R}}^{\text{iid}}(\mathcal{F}; x_{1:n}) \leq \widetilde{\mathcal{O}}\left(\frac{b}{n}\sqrt{\sum_{t=1}^n \|x_t\|_2^2}\right),$$

where the $\widetilde{\mathcal{O}}(\cdot)$ notation hides all logarithmic factors. (This bound is thus of the same order as the one from the last question.)

2. (**Growth function and VC-dimension**)

(a) Let $\mathcal{X} = \mathbb{R}^d$ and $\mathcal{F} = \left\{ f_{\theta,b}(x) = \text{sign}\left(\langle x, \theta \rangle + b\right) \mid \theta \in \mathbb{R}^d, b \in \mathbb{R} \right\}$ be the set of d-dimensional linear classifiers. Prove $\text{VCdim}(\mathcal{F}) = d + 1$ by following the two steps below.

i. (4pts) Construct $d + 1$ points $x_1, \ldots, x_{d+1} \in \mathbb{R}^d$ and argue that for any labeling $y_1, \ldots, y_{d+1} \in \{-1, +1\}$, there exists $f \in \mathcal{F}$ such that $f(x_t) = y_t$ for all $t = 1, \ldots, d + 1$.

ii. (6pts) Prove that for any $d + 2$ points $x_1, \ldots, x_{d+2} \in \mathbb{R}^d$, there exists a labeling $y_1, \ldots, y_{d+2} \in \{-1, +1\}$ such that no $f \in \mathcal{F}$ satisfies $f(x_t) = y_t$ for all $t = 1, \ldots, d + 2$. Hint: consider appending 1 to the end of each of the $d + 2$ points: $(x_1, 1), \cdots, (x_{d+2}, 1) \in \mathbb{R}^{d+1}$, and start with the fact that these $d + 2$ points must be linearly dependent (since they live in $\mathbb{R}^{d+1}$).

(b) (5pts) Let $\mathcal{X} = \mathbb{R}$ and $\mathcal{F} = \{f_\theta(x) = \text{sign}(\sin(\theta x)) \mid \theta \in \mathbb{R}\}$. Prove that for any n, if $x_t = 2^{-2t}$, then $\mathcal{F}$ shatters the set $x_{1:n}$, which means $\text{VCdim}(\mathcal{F}) = \infty$. (Hint: for any labeling $y_{1:n}$, consider $\theta = \pi\left(1 + \sum_{i=1}^{n}(1 - y_i)2^{2i-1}\right)$.)

3. (**Covering number**)

   (a) In Proposition 2 of Lecture 3, via a volumetric argument we show that the linear class $\mathcal{F} = \left\{ f_\theta(x) = \langle \theta, x \rangle \mid \theta \in B_p^d \right\}$ for $\mathcal{X} = B_q^d$ and some $p \geq 1$ and $q \geq 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$ has bounded pointwise covering number: $\mathcal{N}(\mathcal{F}, \alpha) \leq \left(\frac{2}{\alpha} + 1\right)^d$ for any $0 \leq \alpha \leq 1$. Follow the two steps below to further show $\mathcal{N}(\mathcal{F}, \alpha) \geq \left(\frac{1}{2\alpha}\right)^d$.

      i. (5pts) Given any pointwise α-cover $\mathcal{H} \subset [-1, +1]^{\mathcal{X}}$, construct a pointwise 2α-cover $\mathcal{H}' \subset \mathcal{F}$ so that $|\mathcal{H}'| \leq |\mathcal{H}|$ (note that $\mathcal{H}'$ has to be a subset of $\mathcal{F}$).

      ii. (6pts) Prove that if $\mathcal{H}' \subset \mathcal{F}$ is a pointwise 2α-cover of $\mathcal{F}$, then we must have $|\mathcal{H}'| \geq \left(\frac{1}{2\alpha}\right)^d$, which then implies $\mathcal{N}(\mathcal{F}, \alpha) \geq \left(\frac{1}{2\alpha}\right)^d$ as desired. Hint: use a similar volumetric argument.

(b) Let $v_1, \dots, v_d \in B_2^n$ be d points within the n-dimensional ℓ_2-norm unit ball and

$$\mathcal{S} = \left\{ \sum_{i=1}^{d} \beta_i v_i \;\middle|\; \beta_i \geq 0,\ \forall i,\ \text{and} \ \sum_{i=1}^{d} \beta_i \leq B \right\}$$

be the convex hull of these d points scaled by $B > 0$.

i. (5pts) Prove $\mathcal{N}_2(\mathcal{S}, \alpha) \leq \left(\frac{2B}{\sqrt{n}\alpha} + 1 \right)^d$.

ii. Follow the steps below to prove a different covering number bound $\mathcal{N}_2(\mathcal{S}, \alpha) \le d^{\frac{B^2}{n\alpha^2}}$.

A. (4pts) For any $v = \sum_{i=1}^d \beta_i v_i \in \mathcal{S}$, let $\beta = (\beta_1, \dots, \beta_d)$ and define m i.i.d. random variables $u_1, \dots, u_m$, each of which is $\|\beta\|_1 v_i$ with probability $\beta_i / \|\beta\|_1$ for $i = 1, \dots, d$. Prove that the mean of these random variables is v and the variance of $u = \frac{1}{m}\sum_{j=1}^m u_j$ is bounded as:

$$\mathbb{E}\left[\|u - v\|_2^2\right] \le \frac{\|\beta\|_1^2}{m}.$$

B. (7pts) Prove that the following is an α-cover of $\mathcal{S}$ with respect to ℓ_2-norm:

$$\mathcal{S}' = \left\{ \frac{B}{M}\sum_{i=1}^d m_i v_i \;\middle|\; \text{each } m_i \text{ is a nonnegative integer and } \sum_{i=1}^d m_i \le M \right\}$$

where $M = \frac{B^2}{n\alpha^2}$. (The statement $\mathcal{N}_2(\mathcal{S}, \alpha) \le d^{\frac{B^2}{n\alpha^2}}$ then follows immediately.)