Theoretical Machine Learning
Homework 1

Instructor: Haipeng Luo

This homework is due on **9/27, 11:59pm**. See course website for more instructions on finishing and submitting your homework as well as late policy.

1. (**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 $\text{VCdim}(\mathcal{F}) = d + 1$ following the two steps below.

      i. (3pts) 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. (5pts) 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: use the fact that $m+1$ points in an m-dimensional space must be linearly dependent.)

   (b) For $k = 1, \ldots, M$, let r_k be some positive integer and $\mathcal{F}_k : \{-1, +1\}^{r_k} \to \{-1, +1\}$ be some function class with growth function $\Pi_{\mathcal{F}_k}$ and VC-dimension d_k. Further define a vector-valued function class mapping from $\{-1, +1\}^{r_k}$ to $\{-1, +1\}^{r_{k+1}}$ as

$$\mathcal{H}_k = \left\{ h(x) = (f_1(x), \ldots, f_{r_{k+1}}(x)) \;\middle|\; f_1, \ldots, f_{r_{k+1}} \in \mathcal{F}_k \right\}.$$

   Define $r_{M+1} = 1$. Then the following class represents an M-layer feedforward neural net

$$\mathcal{F} = \left\{ h_M \circ \cdots \circ h_1 : \{-1, +1\}^{r_1} \to \{-1, +1\} \;\middle|\; h_1 \in \mathcal{H}_1, \ldots, h_M \in \mathcal{H}_M \right\}$$

   where $\circ$ represents function composition (try to draw a picture to help understand the notation).

      i. (4pts) Prove that the growth function of $\mathcal{F}$ is bounded as

$$\Pi_{\mathcal{F}}(n) \leq \prod_{k=1}^{M} \left(\Pi_{\mathcal{F}_k}(n)\right)^{r_{k+1}}.$$

      ii. (3pts) Let $d = \sum_{k=1}^{M} r_{k+1} d_k$. Prove $\Pi_{\mathcal{F}}(n) \leq (en)^d$.

      iii. (3pts) Further show $\text{VCdim}(\mathcal{F}) = \mathcal{O}(d \ln d)$. (Hint: you might find the inequality $1 + \ln x \leq 2\sqrt{x}$ useful.)

(c) (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 $\theta = \pi\left(1 + \sum_{i=1}^n (1-y_i)2^{2i-1}\right)$.)

(d) (8pts) Suppose $\mathcal{X}$ is an infinite set and $\mathcal{F} \subset \{-1,+1\}^{\mathcal{X}}$ is some function class with infinite VC-dimension. With ℓ being the 0-1 loss, prove $\mathcal{V}^{\text{iid}}(\mathcal{F}, n) \geq 1/4$ for any n. (Hint: read the no free lunch theorem proof again.)

2. (**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 shows $\mathcal{N}(\mathcal{F}, \alpha) \geq \left(\frac{1}{2\alpha}\right)^d$. (Hint: use a similar volumetric argument.)

(b) Recall that a function $f : \mathcal{X} \to \mathbb{R}$ is G-Lipschitz if $|f(x) - f(x')| \leq G|x - x'|$ for any two points $x, x' \in \mathcal{X}$.

i. (5pts) Define $\mathcal{F} = \left\{f \in [-1,+1]^{\mathbb{R}} \mid f \text{ is } G\text{-Lipschitz}\right\}$. For any $G > 0$, construct n points $x_1, \ldots, x_n \in \mathbb{R}$ such that $\mathcal{N}_\infty(\mathcal{F}|_{x_{1:n}}, \alpha) \geq \left(\frac{1}{\alpha}\right)^n$.

ii. (8pts) Let $\mathcal{X} = [0,1]$ and $\mathcal{F} = \left\{f \in [-1,+1]^{\mathcal{X}} \mid f \text{ is } G\text{-Lipschitz}\right\}$. Prove for any $\alpha \leq \min\{1, G\}$, $\ln \mathcal{N}(\mathcal{F}, \alpha) \leq \mathcal{O}\left(\frac{G}{\alpha}\right)$. (Hint: discretize $[0,1]$ evenly into G/α points and $[-1,+1]$ into $2/\alpha$ points, and only consider piecewise linear functions)

iii. (5pts) More generally, one can show that if $\mathcal{X} = [0,1]^d$ for some dimension d and $\mathcal{F} = \left\{f \in [-1,+1]^{\mathcal{X}} \mid f \text{ is } G\text{-Lipschitz}\right\}$, then $\ln \mathcal{N}(\mathcal{F}, \alpha) \leq \mathcal{O}\left(\left(\frac{G}{\alpha}\right)^d\right)$. Based on this fact, prove $\mathcal{R}^{\text{iid}}(\mathcal{F}) = \mathcal{O}\left(\frac{G}{(d\sqrt{n})^{\frac{2}{d}}}\right)$ for any $d > 2$ using Dudley entropy integral. (You are also encouraged to figure out the case $d = 1$ or 2.)