Problem Set 3

Instructions:

- The topics for this problem set are:
  Unit 4 – Agnostic PAC, Uniform Convergence and the VC Dimension.
  Unit 5 – The Fundamental Theorem of PAC Learning.
- Before you start, make sure you are familiar with the course's Homework Policy.

1. Consider the following hypothesis class.

▶ **Definition 1.** *Let $d \in \mathbb{N}$. A homogeneous halfspace of dimension d is a function $f_w : \mathbb{R}^d \to \{1, -1\}$ such that*

$$f_w(x) = \operatorname{sign}(\langle w, x \rangle),$$

where $w \in \mathbb{R}^d$.[1] *The class of homogeneous halfspaces of dimension d is*

$$\mathcal{H}_d = \{f_w : \ w \in \mathbb{R}^d\}.$$

In this question we will compute the VC dimension of $\mathcal{H}_d$.

(a) Show that there exists a set $A \subseteq \mathbb{R}^d$ such that $|A| = d$ and $\mathcal{H}_d$ shatters A. [5 pts]

(b) Let $B = \{x^{(1)}, \ldots, x^{(d+1)}\} \subseteq \mathbb{R}^d$ be a set of cardinality $d+1$. Show that there exist two disjoint sets $I, J \subseteq [d+1]$ at least one of which is nonempty, and positive scalars $\alpha_1, \ldots, \alpha_{d+1} \in \mathbb{R}$, such that

$$\sum_{i \in I} \alpha_i x^{(i)} = \sum_{j \in J} \alpha_j x^{(j)}.$$ [5 pts]

(c) Prove that $\mathsf{VC}(\mathcal{H}_d) = d$. [12 pts]

Hint: Assume for contradiction that $\mathcal{H}_d$ shatters B, and use the linearity of the inner product.

2. Recall that the growth function of a set F of functions $A \to B$ is

$$\tau_F(m) = \sup_{X \subseteq A, |X| = m} \Big| F|_X \Big|.$$

Let $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ be sets. Let $\mathcal{H}$ be a set of functions $\mathcal{X} \to \mathcal{Y}$, and let $\mathcal{G}$ be a set of functions $\mathcal{Y} \to \mathcal{Z}$. Consider the set of compositions,

$$\mathcal{F} = \left\{ f : \ \mathcal{X} \to \mathcal{Z} \ \middle| \ \exists h \in \mathcal{H} \ \exists g \in \mathcal{G} : \ f(x) = g(h(x)) \right\}.$$

Prove that $\tau_{\mathcal{F}}(m) \leq \tau_{\mathcal{H}}(m) \, \tau_{\mathcal{G}}(m)$. [16 pts]

[1] For all $t \in \mathbb{R}$, $\operatorname{sign}(t) = \mathbb{1}(t \geq 0) - \mathbb{1}(t < 0)$.

3. In this question we will prove the following theorem from Unit 5.

▶ **Theorem 8.** *There exists a constant $c > 0$ as follows. Let $\varepsilon, \delta \in (0, 0.9)$, let $\mathcal{X}$ be a nonempty set, let $\mathcal{D}$ be a distribution over $\mathcal{X} \times \{0,1\}$, let $\mathcal{H}$ be a class of functions $\mathcal{X} \to \{0,1\}$, and assume $\mathsf{VC}(\mathcal{H}) = d < \infty$. Let $S \sim \mathcal{D}^m$, where*

$$m \geq c \cdot \frac{d \ln(d/\varepsilon) + \ln(1/\delta)}{\varepsilon}.$$

Then with probability at least $1 - \delta$, S is an ε-net for $\mathcal{H}$ with respect to distribution $\mathcal{D}$.

(a) Double sampling. Let $S' \sim \mathcal{D}^m$ be an additional sample taken independently of S. Define the following events:

$$E_1 = \{\exists h \in \mathcal{H} :\ L_{\mathcal{D}}(h) > \varepsilon \ \wedge \ L_S(h) = 0\},$$
$$E_2 = \{\exists h \in \mathcal{H} :\ L_{\mathcal{D}}(h) > \varepsilon \ \wedge \ L_S(h) = 0 \ \wedge \ L_{S'}(h) \geq \varepsilon/2\}.$$

Prove that $\mathbb{P}_{S \sim \mathcal{D}^m}[E_1] \leq 2\mathbb{P}_{S \sim \mathcal{D}^m, S' \sim \mathcal{D}^m}[E_2]$. [5 pts]

(b) Symmetrization. Denote $S = (z_1, \ldots, z_m)$ and $S' = (z'_1, \ldots, z'_m)$. Consider two random variables $T = (t_1, \ldots, t_m)$ and $T' = (t'_1, \ldots, t'_m)$ that are generated according to the following process. For each $i \in [m]$, with probability $1/2$ we assign

$$t_i = z_i \quad \text{and} \quad t'_i = z'_i$$

and with probability $1/2$ we make the opposite assignment, namely

$$t_i = z'_i \quad \text{and} \quad t'_i = z_i.$$

The assignment for each index i is chosen independently of the assignments for other indices. Let

$$E_3 = \{\exists h \in \mathcal{H} :\ L_{\mathcal{D}}(h) > \varepsilon \ \wedge \ L_T(h) = 0 \ \wedge \ L_{T'}(h) \geq \varepsilon/2\}.$$

Prove that $\mathbb{P}_{S \sim \mathcal{D}^m, S' \sim \mathcal{D}^m}[E_2] = \mathbb{P}_{T,T'}[E_3]$. [3 pts]

(c) Prove that $\mathbb{P}_{T,T'}[E_3] \leq \tau_{\mathcal{H}}(2m) \cdot 2^{(-\frac{\varepsilon m}{2})}$. [10 pts]

(d) Prove Theorem 8. [10 pts]

Hint: Show that it suffices to take m such that $m \geq c' \cdot \frac{d \ln(m/d) + \ln(1/\delta)}{\varepsilon}$ for some constant $c' > 0$. Next, you may use the following inequality without proof. Let $a \geq 1$ and $b > 0$. Then for all $x \in \mathbb{R}$: $x \geq 4a \ln(2a) + 2b \implies x \geq a \ln(x) + b$.

(e) Prove that $\mathcal{H}$ is PAC learnable with sample complexity $O\left(\frac{d \ln(1/\varepsilon) + \ln(1/\delta)}{\varepsilon}\right)$. Prove this directly using the theorem, the definitions of PAC learning and the definition of an ε-net. [8 pts]

4. Consider the following theorem.

▶ **Definition 2.** *Let $\mathcal{X}$ be a set, let $\mathcal{H}$ be a class of functions $\mathcal{X} \to \{0,1\}$, and let $\mathcal{D}$ be a distribution over $\mathcal{X}$. The VC entropy of $\mathcal{H}$ with respect to $\mathcal{D}$ is*

$$\mathsf{VCEnt}_{\mathcal{D},\mathcal{H}}(m) = \ln\left(\mathbb{E}_{S \sim \mathcal{D}^m}\left[\left|\mathcal{H}|_S\right|\right]\right).$$

▶ **Theorem 3.** *There exists a constant $c > 0$ as follows. Let $\varepsilon, \delta \in (0, 0.9)$, let $\mathcal{X}$ be a nonempty set, let $\mathcal{D}$ be a distribution over $\mathcal{X} \times \{0,1\}$, let $\mathcal{H}$ be a class of functions $\mathcal{X} \to \{0,1\}$. Let $\mathcal{D}_x$ be the marginal distribution of $\mathcal{D}$ over $\mathcal{X}$. Assume that $m \in \mathbb{N}$ satisfies*

$$m \geq c \cdot \frac{\mathsf{VCEnt}_{\mathcal{D}_x,\mathcal{H}}(2m) + \ln(1/\delta)}{\varepsilon^2},$$

and let $S \sim \mathcal{D}^m$. Then with probability at least $1 - \delta$, S is an ε-representative sample for $\mathcal{H}$ with respect to distribution $\mathcal{D}$ and the 0-1 loss function.

(a) Prove Theorem 3. [16 pts]

Hint: Carefully read the proof of Theorem 9 in Unit 5.

(b) As a comparison between Theorem 3 and Theorem 9 in Unit 5, give an example where the guarantee provided by Theorem 3 is better than the guarantee provided by Theorem 9. Specifically, for $\varepsilon = \delta = \frac{1}{3}$, present a class $\mathcal{H}$ of functions $\mathcal{X} \to \{0,1\}$, a distribution $\mathcal{D}$ over $\mathcal{X} \times \{0,1\}$, and a number $m \in \mathbb{N}$ such the following two conditions hold:

- $\mathsf{VC}(\mathcal{H}) = \infty$, so Theorem 9 does not guarantee that a finite sample will be ε-representative.
- Theorem 3 implies that with probability at least $1-\delta$, $S \sim \mathcal{D}^m$ is ε-representative.

[10 pts]