COMS 4773 Fall 2026 Problem Set 4

Problem 1. Prove the following theorem.

Theorem 1. *There is an absolute constant $C > 0$ such that the following holds. Let $\mathcal{H} \subseteq \{-1,1\}^{\mathcal{X}}$ be a hypothesis class over $\mathcal{X}$, let P be a probability distribution over $\mathcal{X} \times \{-1,1\}$, and let $\mathbf{S} \sim P^n$. For every $\theta > 0$ and $\delta \in (0,1)$, with probability at least $1-\delta$, every $f \in \operatorname{conv}(\mathcal{H})$ with $\mathsf{err}_\theta(f;\mathbf{S}) = 0$ satisfies*

$$\mathsf{err}(\operatorname{sign} \circ f; P) \leq C \cdot \left(\frac{\log(n) \cdot \log(|\mathcal{H}|)}{\theta^2 n} + \frac{\log(1/\delta)}{n} \right).$$

The improvement relative to what was shown in lecture is that the dependence on n is $O(\log(n)/n)$, rather than $O(\sqrt{\log(n)/n})$. But it only applies to $f \in \operatorname{conv}(\mathcal{H})$ with $\mathsf{err}_\theta(f;\mathbf{S}) = 0$, rather than all $f \in \operatorname{conv}(\mathcal{H})$.

Problem 2. The following theorem can be proved using the probabilistic method:

Theorem 2. *There is a constant $C > 0$ such that the following holds. For any $(x_1, \ldots, x_n) \in \mathcal{X}^n$, any $K \in \mathbb{N}$, any $\mathcal{H} \subseteq \{-1, 1\}^{\mathcal{X}}$, and any $f \in \mathrm{conv}(\mathcal{H})$, there exists $g_f \in \mathrm{avg}_K(\mathcal{H})$ such that*

$$\max_{i \in [n]} |f(x_i) - g_f(x_i)| \leq C \sqrt{\frac{\log(n)}{K}}.$$

Prove the theorem without using the probabilistic method.
Hint: Use the proof of Von Neumann's minimax theorem via Hedge.

Problem 3. In this problem, you will prove a sample complexity lower bound for $(\epsilon, 1/4)$-PAC learning $\mathcal{C} := \mathsf{HLTF}_{\mathbb{R}^n}$ under a specific marginal distribution over $\mathbb{R}^n$, namely the uniform distribution on the unit sphere $S^{n-1} := \{x \in \mathbb{R}^n \mid \|x\| = 1\}$ (which we denote by $\text{Uniform}(S^{n-1})$). Here is the lower bound theorem that you will prove.

Theorem 3. *There exist constants $C, C' > 0$ such that the following holds. For any (deterministic) learning algorithm A, and any $\epsilon \in (0, 1/C)$, there is a probability distribution P over $\mathbb{R}^n \times \{0,1\}$ that satisfies $\mathsf{HLTF}_{\mathbb{R}^n}$-realizability and for which the marginal distribution of P over $\mathbb{R}^n$ is $P_{\mathbf{x}} = \text{Uniform}(S^{d-1})$ such that, if A receives an iid sample of size-m with*

$$m < \frac{n-1}{C'\epsilon},$$

then with probability at least $1/4$, the output $\mathbf{h}$ of A satisfies $\mathsf{err}(\mathbf{h}; P) > \epsilon$.

(a) Let $\mathcal{F}$ be the set of all $\{0,1\}$-valued functions on $\mathbb{R}^n$. Define the function $\mathsf{d} \colon \mathcal{F} \times \mathcal{F} \to \mathbb{R}$

$$\mathsf{d}(f, g) := \Pr(f(\mathbf{u}) \neq g(\mathbf{u}))$$

where $\mathbf{u} \sim \text{Uniform}(S^{n-1})$. Show that $(\mathcal{F}, \mathsf{d})$ satisfies all the requirements to be a metric space, except that $\mathsf{d}(f, g) = 0$ does not have to imply $f = g$. (Such a space is called a *pseudometric space*, and d is called a *pseudometric*.)

(b) For any fixed size-m sample $T = (x_1, \ldots, x_m) \in (S^{n-1})^m$ and any $c \in \mathcal{C}$, define $T^{(c)} := ((x_1, c(x_1)), \ldots, (x_m, c(x_m)))$ to be corresponding sample labeled by c. Suppose A is a (deterministic) learning algorithm that $(\epsilon, 1/4)$-learns $\mathcal{C}$ from any $\mathcal{C}$-realizable distribution P for which $P_{\mathbf{x}} = \text{Uniform}(S^{n-1})$, and suppose $G \subseteq \mathcal{C}$ is a 2ϵ-packing of $(\mathcal{F}, \mathsf{d})$. Show that for $\mathbf{T} := (\mathbf{x}_1, \ldots, \mathbf{x}_m) \sim \text{Uniform}(S^{n-1})^m$,

$$\frac{3}{4}|G| \leq \mathrm{E}\left[\sum_{g \in G} \mathbb{1}\left\{\mathsf{d}(A(\mathbf{T}^{(g)}), g) \leq \epsilon\right\}\right] \leq \mathrm{E}\Big[|\mathcal{C}[\mathbf{T}]|\Big].$$

Note that A need not be a proper learner.

(c) Here are two useful facts about $\mathsf{HLTF}_{\mathbb{R}^n}$.

Fact 1. *For any $T = (x_1, \ldots, x_m) \in (\mathbb{R}^n)^m$,*

$$|\mathsf{HLTF}_{\mathbb{R}^n}[T]| \leq 2\sum_{k=0}^{n-1} \binom{m-1}{k}.$$

Fact 2. *For any $\varepsilon \in (0, 1/4)$, there exists $G \subseteq \mathsf{HLTF}_{\mathbb{R}^n}$ with*

$$|G| \geq \frac{1}{\sqrt{n}}\left(\frac{1}{2\pi\varepsilon}\right)^{n-1}$$

such that G is a 2ε-packing of $(\mathcal{F}, \mathsf{d})$.

Use Facts 1 and 2 with the previous parts to finish the proof of Theorem 3.