EE5604: Intro to Statistical Learning Theory **2024**

Exam 2: October 2024

Instructor: Shashank Vatedka

Instructions: *This is a closed-book exam. You are not permitted to refer to any material or discuss the problem with anyone. Malpractice will be severely punished. Please mention your ROLL Number and name clearly in the answer sheet.*

Justify all your statements clearly. You may use any result proved in class (but clearly state which results you are using), but everything else needs to be proved.

Question 2.1 (3pts). Prove that the empirical Rademacher complexity satisfies the bounded differences property.

Question 2.2. Find the VC dimension for the following hypothesis classes: (5pts each)

1. $\mathcal{X} = \mathbb{R}^2$,
$$\mathcal{H} = \{h(\underline{x}) = 1_{\{\underline{x} \in \mathcal{A}\}} : \mathcal{A} \text{ is an axis-aligned rectangle}\}$$
2. $\mathcal{X} = \mathbb{R}$
$$\mathcal{H} = \{h(x) = 1_{\{\sum_{i=0}^{k} a_i x^i \geq 0\}} : \underline{a} \in \mathbb{R}^{k+1}\}$$
3. $\mathcal{X} = \mathbb{R}$
$$\mathcal{H} = \{h(x) = 1_{\{\sin(\omega x) \geq 0\}} : \omega \in \mathbb{R}\}$$
Hint: Show that for every positive integer m, the set $\{2, 2^2, \ldots, 2^m\}$ is shattered by $\mathcal{H}$

Question 2.3. Shyam is an aspiring machine learning researcher, and is trying to devise a new classification algorithm for a problem with $\mathcal{X} = \mathbb{R}$. As a first step, he wants to compute the VC dimension of the corresponding hypothesis class $\mathcal{H}$. The class $\mathcal{H}$ is some collection of functions that are defined by d free variables/parameters, and each variable/parameter can take values in $\{-1, 0, 1\}$.

He is able to show that the set $\{1, 2, 3, 5\}$ is shattered by $\mathcal{H}$, but that the set $\{1, 2, 3, 4, 5\}$ is not shattered by $\mathcal{H}$.

1. With the above information, what can he conclude about the VC dimension of $\mathcal{H}$? (3pts)
2. Can he conclude that $\mathcal{H}$ is agnostic PAC learnable? Why/Why not? (2pts)
3. Suppose instead that the parameters can take values in $\{-1, 1\}$. What can he conclude about the VC dimension of this class? (3pts)
4. He then observes that computing the ERM is computationally not feasible, and instead defines a new hypothesis class $\mathcal{H}'$ which is similar to the above, but now each variable/parameter can take values in the interval $[-1, 1]$. Using only the information provided above, can you conclude that $\mathcal{H}'$ is agnostic PAC learnable? Why/why not? (2pts)
5. Suppose that he instead modified $\mathcal{H}$ so that there are countably infinitely many parameters, and each parameter can take values in $\{-1, 0, 1\}$. Is it possible to conclude that this is PAC learnable? Is it possible to conclude that it is nonuniformly PAC learnable? Justify. (2+2pts)

Question 2.4 (5pts). Consider any learning algorithm that takes training set S of size n as input and outputs $h^{(S)}$, and consider any distribution p_{XY}, loss function bounded within $[0, 1]$. Prove that the following statements are equivalent:

- For every $\epsilon, \delta > 0$, there exists $m(\epsilon, \delta)$ such that for all $n > m(\epsilon, \delta)$,

$$\Pr[R(h^{(S)}) > \epsilon] < \delta$$

-

$$\lim_{n \to \infty} \mathbb{E}[R(h^{(S)})] = 0$$

where the probability and expectation is over the random S, consisting of n iid sampled drawn according to p_{XY}.