CMU 10-715: Homework 5
VC Dimension and PAC Learnability
DUE: Oct. 19, 2020, 11:59 PM.

Instructions:

- **Collaboration policy:** Collaboration on solving the homework is allowed, after you have thought about the problems on your own. It is also OK to get clarification (but not solutions) from books, again after you have thought about the problems on your own. Please don't search for answers on the web, previous years' homeworks, etc. (please ask the TAs if you are not sure if you can use a particular reference). There are two requirements: first, cite your collaborators fully and completely (e.g., "Alice explained to me what is asked in Question 4.3"). Second, write your solution *independently*: close the book and all of your notes, and send collaborators out of the room, so that the solution comes from you only.

- **Submitting your work:** Assignments should be submitted as PDFs using Gradescope unless explicitly stated otherwise. Each derivation/proof should be completed on a separate page. Submissions can be handwritten, but should be labeled and clearly legible. Else, submission can be written in LaTeX.

- **Late days:** For each homework you get three late days to be used only when anything urgent comes up. No points will be deducted for using these late days. We will consider an honor system where we will rely on you to use the late days appropriately.

1 Finite VC Dimension and PAC Learnability [100]

Consider binary classification, with $\mathcal{X} = \mathbb{R}^d$ and $\mathcal{Y} = \{-1, 1\}$.

For a hypothesis class $\mathcal{H}$, the *restriction* of $\mathcal{H}$ onto $X_n = \{\mathbf{x}_1, \mathbf{x}_2 \ldots, \mathbf{x}_n\}$ is the set of label vectors that can be generated by hypotheses in $\mathcal{H}$, defined as

$$\mathcal{H}_{X_n} = \{h(\mathbf{x}_1), h(\mathbf{x}_2), \ldots, h(\mathbf{x}_n) \mid h \in \mathcal{H}\}.$$

The *growth function* of the hypothesis class $\mathcal{H}$, $\tau_{\mathcal{H}} : \mathbb{N} \to \mathbb{N}$, is defined as:

$$\tau_{\mathcal{H}}(n) = \max_{X_n \subseteq \mathcal{X}:\, |X_n| = n} |\mathcal{H}_{X_n}|.$$

a) (50pts) Recall **Sauer's lemma** seen in class: Let $\mathcal{H}$ be a hypothesis class with $\mathrm{VCdim}(\mathcal{H}) = d < \infty$, then for all $n \in \mathbb{N}$,

$$\tau_{\mathcal{H}}(n) \leq \sum_{i=0}^{d} \binom{n}{i}.$$

Show this Sauer's Lemma implication, that when $n > d + 1$:

$$\tau_{\mathcal{H}}(n) \leq \left(\frac{en}{d}\right)^d$$

where e is the Euler's constant.

b) (50pts) Now consider the following important lemma: Let $\mathcal{H}$ be a hypothesis class and let $\tau_{\mathcal{H}}$ be its growth function. Then for every distribution $\mathcal{D}$ on $\mathcal{X} \times \mathcal{Y}$ and $\delta \in (0, 1)$, with probability of at least $1 - \delta$ over the choice of a sample $S \sim \mathcal{D}^n$, we have

$$|R_{(D)}(h) - R_{(UnifS)}(h)| \leq \frac{4 + \sqrt{\log(\tau_{\mathcal{H}}(2n))}}{\delta \sqrt{2n}}.$$

Using your upper bound for the growth function and the above lemma, you will prove a part of the fundamental theorem of statistical learning. Show that when the hypothesis class $\mathcal{H}$ has finite VC dimension then it is agnostic PAC learnable. In doing so, derive an upper bound on the sample complexity. Note that this bound can be loose as compared to the sample complexity bound in the (agnostic) fundamental theorem of statistical learning theory.