Homework 1
APPM 4490/5490 Theory of Machine Learning, Spring 2024

Due date: Friday, Jan 26 '24, before noon, via paper or via Gradescope **Instructor**: Prof. Becker
Revision date: 1/20/2024

Theme: Introduction, PAC learning, specialized PAC analysis (finite classes, axis-aligned rectangles)

Instructions Collaboration with your fellow students is OK and in fact recommended, although direct copying is not allowed. The internet is allowed for basic tasks (e.g., looking up definitions on wikipedia) but it is not permissible to search for proofs or to *post* requests for help on forums such as `http://math.stackexchange.com/` or to look at solution manuals. Please write down the names of the students that you worked with.

An arbitrary subset of these questions will be graded.

Reading You are responsible for chapters 1, 2, 3 and up to 4.1 in "Understanding Machine Learning" by Shai Shalev-Shwartz and Shai Ben-David (2014, Cambridge University Press). Note: you can buy the book for about $45 on Amazon (or less for an e-book), and the authors host a free PDF copy on their website (but note that this PDF has different page numbers).

Problem 1: Problem 2.1 in Shalev-Shwartz and Ben-David on how thresholded polynomials can memorize a dataset. Given a training set $S = \{(\mathbf{x}_i, f(\mathbf{x}_i))\}_{i=1}^m$ with $\mathcal{X} = \mathbb{R}^d$ and $\mathcal{Y} = \{0, 1\}$, show there exists a polynomial p_S such that $h_S(\mathbf{x}) = 1$ iff $p_S(\mathbf{x}) \geq 0$ where

$$h_S(\mathbf{x}) = \begin{cases} f(\mathbf{x}_i) & \text{if } \exists i \in [m] \text{ s.t. } \mathbf{x}_i = \mathbf{x} \\ 0 & \text{otherwise} \end{cases}.$$

The point of this problem is that the class of thresholded polynomials seems like a quite reasonable class that isn't exotic or unreasonably large, yet even such a reasonable class can memorize data.

Problem 2: Problem 2.3 in Shalev-Shwartz and Ben-David on axis aligned rectangles, and assuming realizability. This has 4 parts. Parts 3 and 4 do not have to be done rigorously. Throughout, assume the distribution $\mathcal{D}$ is a continuous probability distribution.

1) (Show that `A` is an ERM, where `A` returns the smallest rectangle enclosing all positive examples).
2) (Show that if $m \geq 4\log(4/\delta)/\epsilon$ then with probability at least $1-\delta$ it returns a hypothesis h_S with error $L_{(D,f)}(h_S) \leq \epsilon$... following the book's hint with R^*).
   a) (Show $R(S) \subseteq R^*$ [with probability 1])
   b) (Show that if S contains positive examples in all of the rectangles $R_1, \ldots, R_4$ [as defined in the book] then the hypothesis returned by `A` has error at most ϵ)
   c) (For each $i \in [4]$, bound the probability S does not contain an example from R_i)
   d) (Conclude the argument via the union bound)
3) (Going from $\mathbb{R}^2$ to $\mathbb{R}^d$)
4) (Show runtime of `A` is [bounded by a] polynomial in $d, 1/\epsilon$ and $\log(1/\delta)$)

Optional Problem 2.2 in Shalev-Shwartz and Ben-David. This is easy yet could be confusing (e.g., the proof is almost trivial).