MCS 548 – Mathematical Theory of Artificial Intelligence
Fall 2020
Problem Set 1

Lev Reyzin

Due: 9/25/20 at the beginning of class

Instructions: Atop your problem set, please write your name and list your collaborators. You may consult outside references, but cite all the resources used (e.g. which resources on the internet you consulted). You should not, however, search for answers to these questions. All problems in this assignment require proof.

Problems

1. [10 pts.] What is the VC-dimension of axis-aligned rectangles? Use this VC analysis to straightforwardly conclude that axis-aligned are efficiently PAC learnable.

2. [10 pts.] Let the domain be $\mathcal{R}$ and the concept class $\mathcal{C}_s$ (assume the learner knows s) be the class of concepts defined by unions of s intervals: i.e. c is defined by $a_1 \leq a_2 \leq \ldots \leq a_{2s-1} \leq a_{2s} \in \mathcal{R}$ and $c(x) = 1$ if $x \in [a_1, a_2] \cup [a_3, a_4] \cup \ldots \cup [a_{2s-1}, a_{2s}]$. Show that C_s is efficiently PAC learnable.

3. [10 pts.] Consider the following variant of the PAC model. Given a target function $f : \mathcal{X} \to \{0, 1\}$, let $\mathcal{D}^+$ be the distribution over $\mathcal{X}^+ = \{x \in \mathcal{X} : f(x) = 1\}$ defined as $\mathcal{D}^+(a) = \mathcal{D}(a)/\mathcal{D}(\mathcal{X}^+)$ for $a \in \mathcal{X}^+$. And D^- is the distribution over $\mathcal{X}^-$ (defined analogously). In this model, the learner does not have access to D but is able to draw examples from both $\mathcal{D}^+$ and $\mathcal{D}^-$. A class is learnable in this model if a learner can produce a hypothesis h whose risk is $\leq \epsilon$ on both D^- and D^+ simultaneously. Show that if $\mathcal{H}$ is efficiently learnable in the standard PAC model then $\mathcal{H}$ is also efficiently learnable in this variant.

4. [10 pts.] A k-fold union of hypotheses from a class $\mathcal{C}$ is a collection $c_1, \ldots, c_k \in \mathcal{C}$ that assigns the label $c_1(x) \vee \ldots \vee c_k(x)$ to example x. Find a class $\mathcal{C}$ of some VC dimension d such that the class of k-fold unions of hypotheses from $\mathcal{C}$ has VC dimension greater than $(1 + \epsilon)kd$ for sufficiently large values of k.

5. [10 pts.] Let $f : X \to \{+1, -1\}$ be a classifier, and let $C := \{-f, f\}$. Give as strong upper and lower bounds on $\mathfrak{R}_m(C)$ as you can. (For full credit, give bounds that are asymptotically tight in m for any D.)

6. [10 pts.] Consider modifying the definition of PAC learning by getting rid of the δ parameter and letting ϵ serve as a bound on both the approximation error and the failure probability. In essence, the learner would be asked to produce a hypothesis h_S such that

$$\Pr_{S \sim D^m}[R(h_S) \leq \epsilon] \geq 1 - \epsilon$$

using a sample size polynomial in $1/\epsilon$, and the dependence on the other parameters would remain unchanged. Does this redefinition change which classes of functions are PAC learnable?