MCS 548 – Mathematical Theory of Artificial Intelligence

Fall 2018

Problem Set 1

Lev Reyzin

Due: 10/5/18 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. Let the domain be $\mathcal{R}$ and the concept $\mathcal{C}_s$ 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 (in the realizable setting). For both parts, assume the learner knows s (and can choose a sample size m as a function of s, ϵ, and δ). Make sure to argue your learner runs in time polynomial in $s, 1/\epsilon$, and $1/\delta$. Extra credit: show that C_s is efficiently learnable in the agnostic PAC model.

2. Recall that a Boolean literal is either a variable $x_i, i \in [1 \ldots n]$ or its negation $\bar{x}_i$. Give a membership and equivalence query algorithm for efficient exact learning of conjunctions of at most n Boolean literals. Are both equivalence and membership queries necessary for efficient exact learning? If not, do equivalence queries alone suffice? Do membership queries?

3. Consider the following variant of the PAC model. Given a target function $f : \mathcal{X} \rightarrow \{0, 1\}$, 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. 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. Give an explicit class $\mathcal{C}$ of (any) 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.[1] Extra credit will be given for exhibiting a class with growth rate of $\omega(kd)$.

5. Let $f : X \rightarrow \{+1, -1\}$ 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.)

References

[1] Anselm Blumer, Andrzej Ehrenfeucht, David Haussler, and Manfred K. Warmuth. Learnability and the vapnik-chervonenkis dimension. *J. ACM*, 36(4):929–965, 1989.

[1] An upper bound on the VC-dimension of k-fold union of $2kd \log_2(3k)$ is given by Blumer et al. [1].