CIS 625 PROBLEM SET 2

Professor Kearns
Due November 7, 2021 on Gradescope (code ERP5VV)

Problem 6.
Consider the class of unions of d intervals over the real line from Problem Set 1 Problem 2. Compute the VC dimension of this class exactly.

Problem 7.
For any concept class C, we call the set of inputs $S = \{x_1, \ldots, x_m\}$ a **forcing sequence** for $c \in C$ if c is the only concept in C consistent with the labeled sample $\{\langle x_1, c(x_1)\rangle, \ldots, \langle x_m, c(x_m)\rangle\}$. In other words, no other concept in C gives the same labeling on S that c does, and thus any consistent learning algorithm is "forced" to choose c on this labeled sample. The forcing dimension of C, denoted $\mathrm{FD}(C)$, is the smallest natural number d such that every $c \in C$ has a forcing sequence of at most d points.

a) Let C be the class of conjunctions over n boolean variables and their negations. Show that $\mathrm{FD}(C) = n + 2$.
b) Show that there is a finite class C such that $\mathrm{FD}(C) = |C| - 1$ and $\mathrm{VC}(C) = 1$.
c) Show that there is a class C for which $\mathrm{FD}(C) < \mathrm{VC}(C)$. Here $\mathrm{VC}(C)$ is the VC dimension of C.

Problem 8.
In lecture, we sketched the proof showing that for any class C with VC dimension d, there is a distribution P on which any PAC learning algorithm for C requires at least on the order of d/ε samples. Prove this result carefully and rigorously.

Problem 9. Let C be a concept class. We define the class of **k-fold disjunctions** over C, denoted $C[k]$, as the class of all concepts of the form $c_1 \vee c_2 \vee \ldots \vee c_k$, where each $c_i \in C$. As an example, what we have been calling k-term DNF is simply the class of k-fold disjunctions over conjunctions. Show that if C has VC dimension d, the VC dimension of $C[k]$ is $O\left(d \times k \times \log(k)\right)$.

Problem 10. In lecture we showed that if, given a sample S of some $c \in C$, we could find a hypothesis h (not necessarily in C) whose length was sublinear in $m = |S|$, we could PAC learn C by such hypotheses, provided m is sufficiently large. In this problem you are asked to prove the converse. Suppose we have a PAC learning algorithm L for C that outputs hypotheses of some unknown form. Then use L to construct an algorithm that, given a sample S of some $c \in S$, outputs an efficiently evaluatable hypothesis h (whose form you may determine) that is consistent with S, and has size sublinear in m. *Hint: design a simulation of L in which the accuracy parameter depends on the degree of the polynomial running time of L.*

Problem 11.
Extra Credit: Show that if concept class C is PAC learnable by concept class H, then as long as H contains the always-1 concept (i.e. every $x \in X$ is a positive example) and the always-0 concept (every $x \in X$ is a negative example), then C is PAC learnable by H by a deterministic algorithm (i.e. an algo with no internal randomization).