15-859(B) Machine Learning Theory

Homework # 6 **Due: April 26, 2006**

Groundrules: Same as before. You should work on the exercises by yourself but may work with others on the problems (just write down who you worked with). Also if you use material from outside sources, say where you got it.

Exercises:

1. **DFAs.** A *distinguishing sequence* for a DFA is a sequence of actions such that the observations produced from these actions uniquely determines the starting state. I.e., a sequence h such that if $q \neq q'$ then $\text{obs}(q, h) \neq \text{obs}(q', h)$.

   (a) Describe a strongly-connected DFA that has no distinguishing sequence. Note that the definition of "$q \neq q'$" is that there must exist a sequence $h_{qq'}$ such that $\text{obs}(q, h_{qq'}) \neq \text{obs}(q', h_{qq'})$, it's just that no single h works for all pairs.

   (b) Give a homing sequence for your DFA.

Problems:

2. **Fourier.** Suppose f is a function from $\{0,1\}^n$ to $\{-1,1\}$, and g is an approximation to the fourier representation of f. Specifically, $\langle f, g\rangle = \mathbf{E}_{x\in D}[f(x)g(x)] = 1 - \epsilon$ and $g(x) \in [-1,1]$. Then, as we noticed in class, even if g is not a boolean function, $\Pr_{x\in D}(\text{sign}(g(x)) \neq f(x)) \leq \epsilon$. That is because whenever $\text{sign}(g)$ is incorrect we have $f(x)g(x) \leq 0$. So, we can use $\text{sign}(g)$ as a good approximation to f.

   However, what if we are only able to capture "some" of the fourier representation of f. E.g., $\langle f, g\rangle = 0.1$, or even $\langle f, g\rangle = 0.5$. Then, even though g has a reasonable correlation with f in the fourier sense, the guarantee on $\text{sign}(g)$ is vacuous.

   Give an alternate construction $\bar{g}$ based on g with the property that if $\langle f, g\rangle = 1 - \epsilon$ then $\Pr(\bar{g}(x) \neq f(x)) \leq \epsilon/2$. So, this is now useful for the whole range of $\epsilon < 1$. Again, assume $g(x) \in [-1,1]$ for all x. Hint: you may want $\bar{g}$ to be a probabilistic function.

3. **Sample complexity bounds.** For some learning algorithms, the hypothesis produced can be uniquely described by a small subset of k of the training examples. E.g., if you are learning an interval on the line using the simple algorithm "take the smallest interval that encloses all the positive examples," then the hypothesis can be reconstructed from just the outermost positive examples, so $k = 2$. For conservative Mistake-Bound learning algorithm algorithms, you can reconstruct the hypothesis by just looking at the examples on which a mistake was made, so $k \leq M$, where M is the algorithm's mistake-bound.

Prove a PAC guarantee based on k. Specifically, fixing a description language (reconstruction procedure), so for any set S' of examples we have a well-defined hypothesis $h_{S'}$, show that

$$\Pr_{S \sim D^n}\left(\exists S' \subseteq S, |S'| = k, \text{ such that } h_{S'} \text{ has 0 error on } S - S' \text{ but true error } > \epsilon\right) \leq \delta,$$

so long as

$$n \geq \frac{1}{\epsilon}\left(k \ln n + \epsilon k + \ln \frac{1}{\delta}\right).$$

Hint: Think of S' as a subset of indices, and imagine drawing points in S by drawing those in S' first.

Note the similarity of the form of this bound to VC-dimension and other bounds we have seen. These are often called "compression bounds".