Analysis of Boolean Functions **CMU 18-859S, Spring 2007**

PROBLEM SET 3
Due: Thursday, March 8

Homework policy: I encourage you to try to solve the problems by yourself. However, you may collaborate as long as you do the writeup yourself and list the people you talked with.

Do 5 out 7 problems

1. Total influence of DNFs. Let f be computable by a DNF of width w. Show that $\mathbb{I}(f) \leq 2w$. For extra credit, improve on the constant 2.

2. Unbiased functions can't be *that* correlation-immune. Suppose $f : \{-1,1\}^n \to \{-1,1\}$ is dth order correlation-immune (see Homework #1) but $\mathbf{E}[f] \neq 0$. Show that $d < (2/3)n$. (The example from class, $(x_1 \oplus \cdots \oplus x_{(2/3)n}) \wedge (x_{n/3+1} \oplus \cdots \oplus x_n)$, shows that this is tight.) (Hint: $f^2 \equiv 1$.)

3. Weak learning. A *weak* learner is a learning algorithm that does not work for every accuracy parameter ϵ, only for *some* $\epsilon < \frac{1}{2}$. Specifically, we say A γ-*weak-learns* a class if for target function f, its hypothesis h satisfies $\mathbf{E}[fh] \geq \gamma$ (with probability at least $1 - \delta$).

Show that if f is computable by a size-s DNF then there is some $U \subseteq [n]$ with $|U| \leq \log_2(s) + O(1)$ such that $|\hat{f}(U)| \geq \Omega(1/s)$.

(Given this, one can of course $\Omega(1/s)$-weak-learn size-s DNF in poly(s, n) time using membership queries. This is the beginning of Jackson's algorithm.)

4. ϵ-biased sets. Let $\mathcal{R} \subset \{-1,1\}^n$. We say that $\mathcal{R}$ is an ϵ-*biased set* if

$$\left| \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \mathcal{R}} [\boldsymbol{x}_S] \right| \leq \epsilon$$

for every $\emptyset \neq S \subseteq [n]$; here $\boldsymbol{x} \sim \mathcal{R}$ means that $\boldsymbol{x}$ is drawn uniformly at random from $\mathcal{R}$. We say that $\mathcal{R}$ is *efficiently constructible* if there is an algorithm which, on input ϵ and n, writes down all strings in $\mathcal{R}$ in deterministic time $\mathrm{poly}(|\mathcal{R}|, n)$. Later in the course we will show efficiently constructible ϵ-biased sets of cardinality $(n/\epsilon)^2$.

(a) Assume the existence of such efficiently constructible ϵ-biased sets. Given any $S \subseteq [n]$ and query access to some $f : \{-1,1\}^n \to \{-1,1\}$, show how to *deterministically* estimate $\hat{f}(S)$ to within $\pm\epsilon$ in time $\mathrm{poly}(\|\hat{f}\|_1, n, 1/\epsilon)$. You may assume the algorithm knows $\|\hat{f}\|_1$.

(b) In analyzing the spectral norm of DNF in class, we showed that if $(\boldsymbol{I}, \boldsymbol{x})$ is a random restriction, then $\mathbf{E}[\|\widehat{f_{\boldsymbol{x}\to\bar{\boldsymbol{I}}}}\|_1] \leq \|\hat{f}\|_1$. Show the following much stronger result: For *any* restriction $f_{x\to\bar{I}}$ of f, $\|\widehat{f_{x\to\bar{I}}}\|_1 \leq \|\hat{f}\|_1$. Conclude that for any (I, x) and any $S \subseteq I$ we can *deterministically* estimate $F_{S\subseteq I}(x)$ to within $\pm\epsilon$ using queries to f and time $\mathrm{poly}(\|\hat{f}\|_1, n, 1/\epsilon)$.

(With a little bit more work one can similarly estimate $\mathbf{E}_{\boldsymbol{x}}[F_{S\subseteq I}(\boldsymbol{x})]$ for any S and I; this yields a deterministic version of the Goldreich-Levin algorithm running in time $\mathrm{poly}(\|\hat{f}\|_1, n, 1/\epsilon)$. In particular, one gets a polynomial-time *deterministic* algorithm that can exactly recover $O(\log n)$-depth decision trees given membership queries.)

5. Bent functions. Compute the maximum possible value of $\|\hat{f}\|_1$ among functions $f : \{-1,1\}^n \to \{-1,1\}$. Exhibit a function achieving this maximum. (For the latter, you may assume n is odd or even if you want; your choice.)

6. The Low Degree Algorithm's hypothesis.

(a) When doing the Low Degree Algorithm with a fixed d and ϵ, for each $|S| \leq d$ we used an independent batch of random examples to estimate $\hat{f}(S)$. Show that one can in fact first draw a single multiset $\mathcal{E}$ of random examples $(x, f(x))$ of cardinality $\mathrm{poly}(n^d, 1/\epsilon) \cdot \log(1/\delta)$, and then with probability at least $1 - \delta$ have that $(\tilde{\hat{f}}(S) - \hat{f}(S))^2 \leq \epsilon/n^d$ for every $|S| \leq d$, where

$$\tilde{\hat{f}}(S) := \operatorname*{avg}_{(x,f(x))\in\mathcal{E}} \{f(x)x_S\}.$$

(b) Show that if we use this version of the Low Degree Algorithm, our final hypothesis $h : \{-1,1\}^n \to \{-1,1\}$ is of the form

$$h(y) = \mathrm{sgn}\left(\sum_{(x,f(x))\in\mathcal{E}} w(\Delta(y,x)) \cdot f(x)\right),$$

where $w : \{0, 1, \ldots, n\} \to \mathcal{R}$ is some function, and Δ denotes Hamming distance. (In other words, the hypothesis on a given y is equal to a weighted vote over all examples seen, where an example's weight depends only on its Hamming distance to y.) Simplify your expression for w as much as you can.

7. Learning via noise sensitivity. Recall the *noise sensitivity of f at ϵ* from Homework #2, $\mathbb{NS}_\epsilon(f)$. Let $\mathcal{C} = \{f : \{-1,1\}^n \to \{-1,1\} : \mathbb{NS}_\alpha(f) \leq \gamma\}$. Show that the class $\mathcal{C}$ can be learned under the uniform distribution from random examples, to accuracy $O(\gamma)$, in time $\mathrm{poly}(n^{1/\alpha}, 1/\gamma)$.

(E.g., the class of functions such that $\mathbb{NS}_\epsilon(f) \leq O(\sqrt{\epsilon})$ is learnable from random examples, to accuracy ϵ, in time $n^{O(1/\epsilon^2)}$. You might try to convince yourself that Majority$_n$ is in this class, assuming $n \gg 1/\epsilon$.)