6.842 Randomness and Computation	April 4, 2012
Homework 2	
Lecturer: Ronitt Rubinfeld	*Due Date: April 21,2012*

Homework guidelines: You may work with other students, as long as (1) they have not yet solved the problem, (2) you write down the names of all other students with which you discussed the problem, and (3) you write up the solution on your own. No points will be deducted, no matter how many people you talk to, as long as you are honest. If you already knew the answer to one of the problems (call these "famous" problems), then let me know that in your solution writeup – it will not affect your score, but will help me in the future. It's ok to look up famous sums and inequalities that help you to solve the problem, but don't look up an entire solution.

The following problems are to be turned in. TURN YOUR SOLUTION IN TO EACH PROBLEM ON A SEPARATE PIECE OF PAPER WITH YOUR NAME ON EACH ONE.

1. Say that f_1, f_2, f_3, mapping from group G to H, are *linear consistent* if there exists a linear function $\phi : G \to H$ (that is $\forall x, y \in G, \phi(x) + \phi(y) = \phi(x + y)$) and $a_1, a_2, a_3 \in H$ such that $a_1 + a_2 = a_3$ and $f_i(x) = \phi(x) + a_i$ for all $x \in G$. A natural choice for a test of linear consistency is to verify that

   $$Pr_{x,y \in_r G}[f_1(x) + f_2(y) \neq f_3(x + y)] \leq \delta$$

   for some small enough choice of δ.

   - Assume G, H are Abelian. Show that f, g, h are linear-consistent iff for every $x, y \in G$ $f(x) + g(y) = h(x + y)$.
   - Let $G = \{+1, -1\}^n$ and $H = \{+1, -1\}$. First note that since $a_i \in \{+1, -1\}$, then linear consistent f_i must be linear functions or "negations" of linear functions. We refer to the union of linear functions and the negations of linear functions as the *affine functions*. In class we expressed the minimum distance of f to a linear function. Express the minimum distance of a function f to an affine function.
   - Show that if f_1, f_2, f_3 satisfy the above test, then for each $i \in \{1, 2, 3\}$, there is an affine function g_i such that $Pr_{x \in_r G}[f_i(x) \neq g_i(x)] \leq \delta$.
   - (Extra credit) Show that there are linear consistent functions g_1, g_2, g_3 such that for $i \in \{1, 2, 3\}$, $Pr_{x \in_r G}[f_i(x) \neq g_i(x)] \leq \frac{1}{2} - \frac{2\gamma}{3}$ where $\gamma = \frac{1}{2} - \delta$.

2. In order to show that we can save randomness in linearity testing, prove the lemma from class: For all $S \neq \emptyset$, then

   $$E[\Pi_{(i,j) \in S} f(x_i) f(x_j) f(x_i x_j)] \leq \max_\alpha |\hat{f}_\alpha|$$

3. *Dictator functions*, also called *projection functions*, are the functions mapping $\{+1, -1\}^n$ to $\{+1, -1\}$ of the form $f(x) = x_i$ for i in $[n]$.

   Consider the following test for whether a function f is a dictator: Given parameter δ, the test chooses $x, y, z \in \{1, -1\}^n$ by first choosing x, y uniformly from $\{1, -1\}^n$, next choosing w by setting each bit w_i to -1 with probability δ and $+1$ with probability $1 - \delta$ (independently for each i), and finally setting z to be $x \circ y \circ w$, where $\circ$ denotes the bitwise multiply operation. Finally, the test accepts if $f(x)f(y)f(z) = 1$ and rejects otherwise.

- Show that the probability that the test accepts is $\frac{1}{2} + \frac{1}{2}\sum_{s\subseteq[n]}(1-2\delta)^{|S|}\hat{f}(S)^3$.
- Show that if f is a dictator function, then f passes with probability at least $1-\delta$.
- Show that if f passes with probability at least $1-\epsilon$ then there is some S such that $\hat{f}(S)$ is at least $1-2\epsilon$ and such that f is ϵ-close to χ_S.
- Why isn't this enough to give a dictator test? (i.e., what nondictators might pass?) Give a simple fix.

4. Show that if there is a PAC learning algorithm for a class C (sample complexity $poly(\log n, 1/\epsilon, 1/\delta)$) then there is a PAC learning algorithm for C with sample complexity dependence on δ (the confidence parameter) that is only $\log 1/\delta$ – i.e., the "new" PAC algorithm should have sample complexity $poly(\log n, 1/\epsilon, \log 1/\delta)$. (It is ok to assume that the learning algorithm is over the uniform distribution on inputs, although the claim is true in general.)

5. - Show that for any monotone function $f : \{+1,-1\}^n \to \{+1,-1\}$, the influence of the i^{th} variable is equal to the value of the Fourier coefficient of $\{i\}$, that is $\mathrm{inf}_i(f) = \hat{f}(\{i\})$.
   - Show that the majority function $f(x) = \mathrm{sign}(\sum_i x_i)$ maximizes the total influence among n-variable monotone functions mapping $\{+1,-1\}^n$ to $\{+1,-1\}$, for n odd.

6. Consider the sample complexity required to learn the class of monotone functions mapping $\{+1,-1\}^n$ to $\{+1,-1\}$ over the uniform distribution (without queries).

   (a) Show that
   $$\sum_{|S|\geq \mathrm{Inf}(f)/\epsilon} \hat{f}(S)^2 \leq C\cdot\epsilon$$
   where $\mathrm{Inf}(f)$ is the influence of f, and C is an absolute constant.

   (b) Show that the class of monotone functions can be learned to accuracy ϵ with $n^{\Theta(\sqrt{n}/\epsilon)} = 2^{\tilde{O}(\sqrt{n}/\epsilon)}$ samples under the uniform distribution.

Useful definitions:

1. For $x = (x_1, \ldots, x_n) \in \{+1,-1\}^n$, $x^{\oplus i}$ is x with the i-th bit flipped, that is,
$$x^{\oplus i} = (x_1, \ldots, x_{i-1}, -x_i, x_{i+1}, \ldots, x_n).$$
The *influence of the i-th variable on* $f : \{+1,-1\}^n \to \{+1,-1\}$ is
$$\mathrm{Inf}_i(f) = \Pr_x\left[f(x) \neq f\left(x^{\oplus i}\right)\right].$$
The *total influence* of f is
$$\mathrm{Inf}(f) = \sum_{i=1}^{n} \mathrm{Inf}_i(f).$$

2. A function $f : \{+1,-1\}^n \to \{+1,-1\}$ is *monotone* if for all $x, y \in \{+1,-1\}^n$ such that $x_i \leq y_i$ for each i, $f(x) \leq f(y)$. Assume that $-1 < +1$.