6.842 Randomness and Computation March 12, 2008

Homework 3

Lecturer: Ronitt Rubinfeld *Due Date: April 2, 2008*

Homework guidelines: Same as for homework 1.

1. In this problem, we 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.

2. Let G be a bipartite graph with n left vertices and n right vertices. We say that G is a *(bipartite)* (α, γ)*-expander* if for any set S of at most αn left vertices, the size of the neighborhood of S is at least $\gamma|S|$.

   We construct an expander G by independently and uniformly choosing D right neighbors for each left vertex.

   (a) Let S be a subset of left vertices of G. Imagine that we add edges outgoing from S one by one. Argue that the probability that a new edge connects S with a right node that was already in the neighborhood of S is at most $D|S|/n$.

   (b) Prove that the probability that the neighborhood of S is smaller than $|S|(D-2)$ is at most $\binom{D|S|}{2|S|}\left(\frac{D|S|}{n}\right)^{2|S|}$.

   (c) Show that for every D, there is a constant $\alpha > 0$ such that the probability that there is a subset of $t \leq \alpha n$ left vertices that has a neighborhood smaller than $(D-2)t$ is at most 4^{-t}.

   Hint: Use the inequality $\binom{n}{k} \leq \left(\frac{n \cdot e}{k}\right)^k$.

   (d) Conclude that G is a bipartite $(\alpha, D-2)$-expander with probability at least $1/2$.

3. In this problem we develop an efficient approximation algorithm for counting the number of satisfying assignments to a DNF formula $C_1 \vee \ldots \vee C_m$.

   (a) Explain why the naive algorithm that uniformly and indepedently picks random assignments and estimates the fraction of those that satisfy the formula is not what we want.

   (b) Let S be a subset of $\{0,1\}^n \times \{1,\ldots,m\}$ that consists of pairs (x,i) such that the assignment x satisfies C_i. Show that one can efficiently compute the exact size of S in time polynomial in n and m.

(c) Show how to uniformly generate a random element from S in time polynomial in n and m.

(d) Let

$$S' = \{(x,i) \mid \text{there is no } (x,j) \in S \text{ s.t. } j < i\}.$$

Why does $|S'|$ equal the number of assignments satisfying the DNF formula? How can one check if an element of S belongs to S' in time polynomial in n and m?

(e) Show that $m \cdot |S'| \geq |S|$, and using this, show that a $(1+\epsilon)$ approximation to the size of S' can be computed in time polynomial in n, m, and $1/\epsilon$. (Hint: Estimate $|S'|/|S|$.)

4. **(Due 04/07)** Let us first introduce a few definitions.

- Let D be a distributions on n bit strings and f be a function on n bit inputs. We say that f is *ϵ-hard on D for size g* if for any Boolean circuit C with at most g gates, and for x chosen according to D, $\Pr[C(x) = f(x)] \leq 1 - \epsilon$.
- Let M be a measure. If for any circuit C of size g, $\mathrm{Adv}_C(M) < \gamma|M|$, we call f *γ-hard-core* on M for size g. We call f *γ-hard-core on S for size g* if it is hard on the characteristic function of S with the same paramters. We call f *γ-hard-core for size g* if f is on the set of all inputs for the same parameter.

Show the following:

(a) Let f be ϵ-hard for size g on the uniform distributions on n-bit strings, and let $0 < \delta < 1$. Then there is a measure M with with $\mu(M) \geq \epsilon$ so that f is δ-hard-core on M for size $\epsilon^2\delta^2 g/4$.

Update: Assume that $\epsilon^2\delta^2 g/4$ is greater than a constant C such that for any i, there is a circuit of size at most $C \cdot i$ that computes majority of i bits.

(b) Let M be a measure such that f is $\delta/2$-hard-core on M for size $g \in (2n, (1/8)(2^n/n)(\epsilon\delta)^2)$, and assume $\mu(M) \geq \epsilon$. Then there is a set S such that f is 2δ-hard-core on S for size g with $|S| \geq \frac{\epsilon \cdot 2^n}{2}$.

Hint 1: Show that the number of circuits on size g is at most

$$(2(2n+g))^{2g} \leq 2^{2ng} \leq \frac{1}{4} \cdot e^{2^n \epsilon^2 \delta^2}.$$

Hint 2: You can use Hoeffding's inequality. Let X_1 to X_n be independent variables such that for each i, there is a_i such that $X_i \in [a_i, a_i + 1]$. Let $S = \sum_{i=1}^n X_i$. It then holds

$$\Pr(S - E[S] \geq nt) \leq e^{-2nt^2}.$$