6.842 Randomness and Computation April 21, 2014

Homework 10

Lecturer: Ronitt Rubinfeld *Due Date: April 28, 2014*

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. You should upload your solution to Stellar as a pdf file.

1. Suppose that C is a PAC-learnable concept class. Show that there exists a polynomial p such that for all concepts $c \in C_n$ of size s, there exists a circuit of size $p(s,n)$ exactly computing c.

2. **(Learning parity from noisy random examples)** Let $c \in \{0,1\}^n$ be an unkown string and assume that we are given access to examples of the form $\langle x, c \cdot x + z \rangle$ where $x \in_R \{0,1\}^n$ and the noise variable $z \in \{0,1\}$ is chosen from a Bernouilli distribution with parameter $\eta \in (0, 1/2)$. Here, $+$ denotes addition modulo 2.

   (a) We assume that the gap $1/2 - \eta$ is a constant. If you are allowed time 2^n, how can we recover c with high probability using poly(n) samples ?

   The goal of the remainder of this problem is to design an algorithm that recovers c with high probability while using time and number of samples that are both *subexponential in n*.

   (b) Let $(x_1, l_1), \ldots, (x_s, l_s)$ be noisy random examples where $l_i := c \cdot x_i + z_i$ for every $i \in [s]$. Show that $l_1 + \ldots + l_s$ is the correct value of $(x_1 + \ldots + x_s) \cdot c$ with probability $\frac{1}{2} + \frac{1}{2}(1 - 2\eta)^s$.

   (c) Let $n = ab$; we will view $[n]$ as a sequence of a blocks each consisting of b consecutive bits. Let V_i be the subspace of $\{0,1\}^{ab}$ consisting of all vectors whose rightmost i blocks have all bits equal to 0. An i-sample of size s is a set of vectors independently and uniformly distributed over V_i. Assume that we are given an i-sample of size s. Show that we can in time $O(s)$ construct an $(i+1)$-sample of size $s - 2^b$ such that each vector in the $(i+1)$-sample can be written as the sum of 2 vectors in the given i-sample.

   (d) Draw $a2^b$ labeled examples; note that they form a 0-sample of size $a2^b$. Use them along with part (c) to construct a $(a-1)$-sample of size at least 2^b. What's the probability that this $(a-1)$-sample contains the first standard basis vector $(1, 0, \ldots, 0)$?

(e) Use parts (b) and (d) to show that we can recover $(1, 0, \ldots, 0) \cdot c$ with probability $\frac{1}{2} + \frac{1}{2}(1 - 2\eta)^{2^{a-1}}$. Note that $(1, 0, \ldots, 0) \cdot c$ is the first bit of c. How can you recover any other fixed bit of c ?

(f) Use the previous parts to get an algorithm that recovers c with high probability in time and number of examples $\mathrm{poly}((\frac{1}{1-2\eta})^{2^a}, 2^b)$. Observe that if the gap $1/2 - \eta$ is a constant, this gives a bound of $2^{O(n/\log n)}$ for some choice of a and b.