Homework 6
Due 2012/11/19

Oded Regev
Courant Institute

1. **Total influence of DNFs:**

   (a) Assume f can be expressed as a DNF of width w (i.e., each clause has at most w literals). Show that $\mathbb{I}(f) \leq 2w$. Hint: no need for Fourier. Open question: improve on the constant 2.

   (b) Deduce that width-w DNFs can be learned from random examples in time $n^{O(w/\varepsilon)}$. We will improve this in class.

2. **Learning functions with low $\|\hat{f}\|_1$:**

   (a) For $f : \{0,1\}^n \to \mathbb{R}$ let $L = \|\hat{f}\|_1 := \sum_S |\hat{f}(S)|$. Show that for any $\varepsilon > 0$, f is ε-concentrated on a set of size at most L^2/ε.

   (b) Deduce that the set of Boolean functions f with $\|\hat{f}\|_1 \leq L$ can be learned in time $\mathrm{poly}(L, \frac{1}{\varepsilon}, n)$ using membership queries.

   (c) Define a *decision tree on parities* as a decision tree where on each node we can branch on an arbitrary parity of variables (as opposed to just a single variable in the usual definition of decision trees). Show that decision trees on parities of size L can be learned in time $\mathrm{poly}(L, \frac{1}{\varepsilon}, n)$ using membership queries.

3. **A variable that is often much smaller than its expectation has high variance:** Show that if X is a nonnegative random variable with $\Pr[X > K] = \delta$ and $\mathbb{E}[X] \geq L > K$ then $\mathbb{E}[X^2] \geq (L-K)^2/\delta$.