6.896 Sublinear Time Algorithms May 1, 2007

Homework 1

Lecturer: Ronitt Rubinfeld *Due Date: May 10, 2007*

This homework set is *optional!*

1. (Property testing of the clusterability of a set of points.) Given a set X of points in any metric space. Assume that one can compute the distance between any pair of points in one step. Say that X is (k, b)-diameter clusterable if X can be partitioned into k subsets (clusters) such that the maximum distance between any pair of points in a cluster is b. Say that X is ϵ-far from (k, b)-diameter clusterable if at least $\epsilon|X|$ points must be deleted from X in order to make it (k, b)-diameter clusterable.

   Show how to distinguish the case when X is (k, b)-diameter clusterable from the case when X is ϵ-far from $(k, 2b)$-diameter clusterable. Your algorithm should run in time polynomial in $k, 1/\epsilon$. It is possible to get an algorithm which runs in time $O((k^2 \log k)/\epsilon)$.

   Note that the [ADPR] result presented in class works only for a specific metric space, whereas this algorithm should work in general.

2. A *vertex cover* V' of a set of edges E' is a set of nodes such that every edge of E' is adjacent to one of the nodes in V'.

   Let $f : V \to \{0, 1\}$ be a labeling of the vertices of a known directed acyclic graph G by 0 and 1. For any pair of nodes x and y, we say that $x \leq_G y$ if there is a path from x to y in G. We say that f is *monotone* if for all $x \leq_G y$, $f(x) \leq f(y)$. The *minimum distance of f to monotone* is the minimum number of nodes that must be relabeled in order to turn f into a monotone function.

   Let E' be the set of violating edges according to f. Show that the minimum distance of f to monotone is equal to the minimum size of a vertex cover of E'.

3. 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 $\gamma = \frac{1}{2} - \delta$.

4. 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 $\inf_i(f) = \hat{f}(\{i\})$.

5. Show that the majority function $f(x) = \text{sign}(\sum_i x_i)$ maximizes the total influence among n-variable monotone functions mapping $\{+1, -1\}^n$ to $\{+1, -1\}$, for n odd.