Randomized Algorithms 2015A – Problem Set 3

Robert Krauthgamer and Moni Naor

Due: January 18, 2015

1. Show that the hash function $h_r : \{0,1\}^n \to \{0,1\}$ mapping $x \mapsto \sum_{i=1}^{n} x_i r_i \pmod 2$, where $\vec{r} \in \{0,1\}^n$ is chosen uniformly at random, is a good one-bit sketch for equality testing.

   Hint: Analyze $\Pr_r[h_r(x) = h_r(y)]$ when $x = y$ and when $x \neq y$.

2. Suppose the set of possible inputs (of size n) is partitioned into three sets called CLOSE, FAR, and UNKNOWN. Suppose that the randomized algorithm A has advantage $\varepsilon > 0$ in distinguishing CLOSE from FAR inputs in the following sense: there is $p = p_{\text{close}} > 0$ such that

   - for every input z in CLOSE, $\Pr[A(z) = 1] \leq p_{\text{close}}$; and
   - for every input z in FAR, $\Pr[A(z) = 1] \geq p_{\text{close}} + \varepsilon$.

   Design algorithm B that uses $m = O(\frac{1}{\varepsilon^2} \log \frac{1}{\delta})$ independent repetitions of A to distinguish between CLOSE and FAR inputs with success probability (in each of the cases) at least $1 - \delta$.

   Hint: Think first about $\delta = 1/4$.

3. Given as input n points $x_1, \ldots, x_n \in [m]^d$ for $m = d = n/10$, show how to determine, within $1 + \varepsilon$ approximation, the radius of the point set under ℓ_2-distance, defined as $r = \min_{i \in [n]} \max_{j \in [n]} \|x_i - x_j\|$.

   You algorithm should be faster than the naive computation that runs in time $O(n^2 d)$, which in our case is $O(n^3)$.

Extra credit:

4. (a) Let $x_1, \ldots, x_n \in \mathbb{R}^d$ and fix a linear map $L : \mathbb{R}^d \to \mathbb{R}^t$ that preserves all pairwise distances within factor $1+\varepsilon$ (i.e., $\|L(x_i - x_j)\| \in (1 \pm \varepsilon)\|x_i - x_j\|$ for all i, j). Prove that the area of every right-angled triangle $\{x_i, x_j, x_k\}$ (i.e., whenever the inner-product $\langle x_j - x_i, x_k - x_i \rangle = 0$) is preserved by L within factor $1 + O(\varepsilon)$.

   Hint: Denote the triangle's sidelengths by $v = x_j - x_i$ and $w = x_k - x_i$, and let $\hat{v}, \hat{w}$ be defined similarly for the image triangle. Then prove that $|\langle \hat{v}, \hat{w} \rangle| \leq O(\varepsilon) \cdot \|\hat{v}\| \cdot \|\hat{w}\|$.

   (b) Show there is a random map $L : \mathbb{R}^d \to \mathbb{R}^t$ for $t = O(\varepsilon^{-2} \log n)$, such that for every n points $y_1, \ldots, y_n \in \mathbb{R}^d$, with high probability, L preserves the area of every triangle $\{y_i, y_j, y_k\}$ within factor $1 + \varepsilon$.

Hint: For every triangle, find an additional point that "breaks" the triangle into two right-angle triangles. Augment the point set with these $O(n^3)$ additional points, and apply the JL-lemma on this augmented point set.