CS 278 Complexity Theory
Problem Set 3

Due: Mar 18, 2020, 11:59 PM PST

Submission on Gradescope. You are encouraged to discuss the problems and solve them in groups. However, the solutions are to be written up alone, **listing all the collaborators**.

1. **Statistical Distance.** (5 points per item) Recall that a distribution $\mathcal{D}$ over $\{0,1\}^n$ is ε-pseudorandom against a class of functions $\mathcal{C}$ if

$$\forall f \in \mathcal{C} : |\operatorname*{\mathbf{Pr}}_{x\sim\mathcal{D}}[f(x)=1] - \operatorname*{\mathbf{Pr}}_{x\sim\mathcal{U}_n}[f(x)=1]| \leq \varepsilon$$

   (where $\mathcal{U}_n$ is the uniform distribution over $\{0,1\}^n$).

   Suppose we wanted a distribution $\mathcal{D}$ that is ε-pseudorandom against **all** functions $f\colon \{0,1\}^n \to \{0,1\}$. If $\mathcal{D}$ is such a distribution, then we say that the statistical distance between $\mathcal{D}$ and $\mathcal{U}_n$ is at most ε.

   (a) Show that for any $\varepsilon > 0$ there exists a distribution $\mathcal{D}$ which is ε-pseudorandom against all functions from $\{0,1\}^n \to \{0,1\}$ but $\mathcal{D} \neq \mathcal{U}_n$.

   (b) Let $\mathrm{supp}(\mathcal{D})$ be the set of strings which are attained with positive probability in $\mathcal{D}$. Show that if $\mathcal{D}$ is ε-pseudorandom against all functions from $\{0,1\}^n \to \{0,1\}$ then $|\mathrm{supp}(\mathcal{D})| \geq (1-\varepsilon)\cdot 2^n$.

   (c) Suppose $G : \{0,1\}^d \to \{0,1\}^n$ is a pseudorandom generator that ε-fools all functions from $\{0,1\}^n$ to $\{0,1\}$. Show that if $\varepsilon < 0.5$, then $d \geq n$.

   (d) Show that if

$$\sum_{y\in\{0,1\}^n} \left| \operatorname*{\mathbf{Pr}}_{x\sim\mathcal{D}}[x=y] - \operatorname*{\mathbf{Pr}}_{x\sim\mathcal{U}_n}[x=y] \right| \leq \varepsilon$$

   then the statistical distance between $\mathcal{D}$ and $\mathcal{U}_n$ is at most ε.

2. **Yao's XOR Lemma and Impagliazzo's Lemma for Formulas.** (10 points per item) In this exercise, we will adjust XOR Lemma and Impagliazzo's Lemma for formulas (i.e., circuits with fan-out 1, or equivalently binary trees with AND, OR, NOT on internal nodes). We denote the size of a formula as the number of leaves in the tree describing it.

   (a) You can use the following fact: there exists a formula of size at most m^6 that computes the Majority of m bits. Adjust the statement of Impagliazzo's lemma for the case of formulas, and give a proof sketch highlighting the differences from the proof we saw in class.

   (b) Let $f : \{0,1\}^n \to \{0,1\}$ be non-constant. Suppose f cannot be computed by formulas of size s, show that $f^{\oplus k}$ cannot be computed by formulas of size sk.

(c) (Extra Credit) Let $f : \{0,1\}^n \to \{0,1\}$, $\varepsilon, \delta \in (0,1)$ with $\varepsilon > 2(1-\delta)^{k/2}$. Suppose H is a δ-dense distribution such that any formula F of size s satisfies

$$\Pr_{x \sim H}[F(x) = f(x)] \leq \frac{1}{2} + \frac{\varepsilon}{2}.$$

Show that for any formula F' of size $sk/2$ on nk variables,

$$\Pr_{x \sim \mathcal{U}_{n \cdot k}}[f^{\oplus k}(x) = F'(x)] \leq \frac{1}{2} + \varepsilon.$$

(Hint: Think of the nk variables as k blocks of n variables each. A formula with at most $sk/2$ leaves in total would have at most $k/2$ blocks with more than s leaves marked by variables from that block.)

3. **Error Correcting Codes with Relative Distance 1/2.** Let $Enc : \{0,1\}^k \to \{0,1\}^n$ be an error correcting code with distance $n/2$. Our goal is to show that $k \leq \log_2(2n)$.

   Identify every codeword $y = Enc(m)$ with a vector $z \in \{-1,1\}^n$, with $z_i = (-1)^{y_i}$.

   (a) (5 points) Show that if the Hamming distance between two codewords $y, y' \in \{0,1\}^n$ is at least $n/2$, then their corresponding $\{-1,1\}^n$ vectors z and z' have non-positive inner product.

   (b) (10 points) Let $v^{(1)}, \ldots, v^{(m)}$ be a collection of m vectors in $\mathbb{R}^n$ with pairwise non-positive inner product (i.e. $\forall i \neq j : \langle v^{(i)}, v^{(j)} \rangle \leq 0$) and positive first coordinates (i.e., $v_1^{(i)} > 0$ for all $i = 1, \ldots, m$). Show that $m \leq n$.

   (c) (5 points) Deduce that any error correcting code with distance $n/2$ has at most $2n$ codewords. In other words, that $k \leq \log_2(2n)$.