CS 278 - Homework 3

Joshua Turcotti

March 16, 2021

1 Statistical Distance

(a) We choose two arbitrary strings in $\{0,1\}^n$. To define a distribution $\mathcal{D}$ assign a probability of $\frac{1-\epsilon}{2^n}$ to the first, a probability of $\frac{1+\epsilon}{2^n}$ to the second, and a probability of $\frac{1}{2^n}$ to every other string. For any function $\{0,1\}^n \to \{0,1\}$, the maximal difference between its probability of taking the value 1 under $\mathcal{D}$ and $\mathcal{U}_n$ is $\frac{\epsilon}{2^{n-1}} < \epsilon$. Thus $\mathcal{D}$ is $\epsilon-$pseudorandom against all functions $\{0,1\}^n \to \{0,1\}$.

(b) Assume for contradiction that at least $\epsilon 2^n$ strings have probability 0 of occuring under $\mathcal{D}$. Let f be the function that is 1 on exactly those strings. Under $\mathcal{D}$, the probability that f takes the value 1 is 0, while under $\mathcal{U}_n$, it is ϵ. Thus $\mathcal{D}$ is not ϵ-pseudorandom against f, so by contradiction of our assumption it must be the case that the support of $\mathcal{D}$ has size at least $(1-\epsilon)2^n$.

(c) Assume $d < n$. Then at most half of the strings in $\{0,1\}^n$ are in the range of G. Let f take the value 1 on all strings $\{0,1\}^n$ that are not in the range of G. We know that under $\mathcal{U}_n$, f has probability at least $1/2$ of taking the value 1, but under $G(\mathcal{U}_d)$, it has probability 0. This contradicts the assumption that G ϵ-fools f, so it must be the case that $d \geq n$.

(d) Choose any $f : \{0,1\}^n \to \{0,1\}$.

$$\begin{aligned}
\left|\Pr_{x\sim\mathcal{D}}[f(x)=1] - \Pr_{x\sim\mathcal{U}_n}[f(x)=1]\right| &= \left|\sum_{y\in\{0,1\}^n} f(y)\Pr_{x\sim\mathcal{D}}[x=y] - \sum_{y\in\{0,1\}^n} f(y)\Pr_{x\sim\mathcal{U}_n}[x=y]\right| \\
&\leq \sum_{y\in\{0,1\}^n} f(y)\left|\Pr_{x\sim\mathcal{D}}[x=y] - \Pr_{x\sim\mathcal{U}_n}[x=y]\right| \\
&\leq \sum_{y\in\{0,1\}^n} \left|\Pr_{x\sim\mathcal{D}}[x=y] - \Pr_{x\sim\mathcal{U}_n}[x=y]\right|
\end{aligned}$$

2 Yao and Impagliazzo for Formulas

(a) The only adjustment to Impagliazzo's lemma that need be made is to replace $S' = \frac{S\epsilon^2}{100n}$ with $S' = \frac{S\epsilon^1 2}{100n^6}$. With this adjustment made, we can walk through every step of the proof presented in class, beginnign by assuming that for all δ-dense distributions H there exists a formula F of size at most S' such that the probability $f(x) = F(x)$ is greater than $\frac{1}{2} + \epsilon$. Phrasing the choice of distribution over formulas F and the δ-dense distributions H to optimize this probability as a zero sum game, the minimax theorem allows us to conclude that there exists a distribution

3 Error Correcting Codes with Relative Distance

(a) For vectors in $\{-1, 1\}^n$, it is clear that their innder product is equal to $c - d$, where c is the number of indices $i \leq n$ at which they agree and d is the number at which they differ. Thus, if $d \geq n/2$ by assumption, then $c - d \leq 0$.

(b) Assume $m > n$. It is impossible for all m vectors to be linearly independent, so, WLOG, assume $v^{(1)}$ lies in the subspace generated by $v^{(2)}, \ldots, v^{(m)}$. We have the identity

$$v^{(1)} = \langle v^{(1)}, v^{(2)} \rangle v^{(2)} + \ldots + \langle v^{(1)}, v^{(m)} \rangle v^{(m)}$$

we can consider only the first coordinates of each vector in this identity, and realize that we have expressed a positive number as a linear combination of positive numbers in which all of the coefficients are non-positive. This is impossible, so we can conclude that $m \leq n$.

(c) We desire an error correcting code with distance $n/2$ in $\{0, 1\}^n$. We showed in part a above that the set of codewords (mapped via $0 \mapsto 1, 1 \mapsto -1$ into $\{-1, 1\}^n$) must have pairwise non-positive inner products, so part b above tells us that there can be at most n codewords that begin with a 0. We can consider the alternate mapping $\{0, 1\}^n \to \{-1, 1\}^n$ that equals the above mapping on all but the first bit, but first peforms $b \mapsto 1 - b$ on the first bit. We note that the result from part a still applies to this mapping, and combining with part b we can conclude that there can be at most n codewords that begin with a 1, as these will be the ones under the new mapping that have positive first coordinates after the mapping. By a union bound, there are at most $2n$ codewords.