Class 8 Exercises

CS250/EE387, Winter 2025

1. In the lecture videos/notes, we saw the "Kautz-Singleton" construction for group testing matrices, and we instantiated it using RS codes. Say that $N = 300$ and $d = 2$ and you want to build a RS-based Kautz-Singleton group testing matrix. How will you choose parameters for q, k? What will your final group testing matrix look like? How many tests does it use? (Note: you may need to come up with a group testing matrix for $N' > N$ items, and then drop some items, since 300 is not a power of a prime).

Solution

Following the Note, let's take $N' = 343 = 7^3$. Then we can choose $q = 7$ and $k = 3$, so that $N' = q^k$. Then we'll just drop 43 of the items to get 300.

Following the construction from the notes: In the lecture notes, we chose the full-length RS code, that is, with $n = q$. In this case, the number of tests is q^2, which is 49. We need to check that $\mathrm{dist}(C) > n\left(\frac{d-1}{d}\right) = 7/2$ for the matrix to be d-disjunct. Fortunately, $\mathrm{dist}(C) = q - k + 1 = 7 - 3 + 1 = 5$, which is indeed larger than $7/2 = 3.5$. So the final matrix is 49×300, where each of the 300 columns are associated with a polynomial of degree at most 2 over $\mathbb{F}_7$, and each of the rows are associated with a pair of numbers (i, j) for $i, j \in \{0, 1, \ldots, 6\}$. The entry indexed by (i, j) and f is 1 if $f(i) = j \mod 7$ and 0 otherwise.

Slightly better by taking a non-full-length RS code: Say we choose an RS code over $\mathbb{F}_7$ of length 5. Then we'd need to check that $\mathrm{dist}(C) > n\left(\frac{d-1}{d}\right) = 5/2 = 2.5$. Fortunately, the distance is $n - k + 1 = 5 - 3 + 1 = 3$, which is indeed larger than 2.5. This approach leads to 35 tests.

Can you do better by looking at different-length RS codes over a different field? Formally, we're looking for q, k so that $q^k \geq 300$, and so that $(2k-1) \cdot q$ is as small as possible. (That's because we need $\mathrm{dist}(C) > n\left(\frac{d-1}{d}\right)$ aka $n - k + 1 > n/2$ aka $n > 2k - 2$, so we can choose $n = 2k - 1$, and then the number of tests is $n \cdot q = (2k-1) \cdot q$, and we want to minimize that.) I get 35 tests in a different way by setting $k = 4, q = 5$, but I haven't optimized among all possibilities for q...

2. In this problem we will adapt the Kautz-Singleton construction from the lecture videos/notes to deal with false negatives and false positives. The set-up is the same: we have N items, at most d of which are positive, and we wish to make T tests. However, now there may be up to E false negatives and E false positives. (Here, a "false positive" is a test that does not contain any positive items but comes up positive anyway; a "false negative" is a test that does contain a positive item but comes up negative).

   (a) Come up with a condition that is similar to d-disjunctness and prove a statement like "if a pooling matrix Φ satisfies [your condition], then Φ can identify up to d positive items, even with up to E false positives and E false negatives. Assume that the false negatives/positives are worst-case.

Solution

A natural condition is the following:

Definition 1. *A matrix $\Phi \in \{0,1\}^{T \times N}$ is (d, E)-disjunct if for any set $\Lambda \subseteq [N]$ of size d,* *and any other $i \in [N] \setminus \Lambda$, there are at least $2E+1$ values of $j \in [T]$ so that $\Phi_{j,i} = 1$ and $\Phi_{j,r} = 0$ for all $r \in \Lambda$.*

Now we'll prove that this definition is enough to identify up to d positive items, even with E false positives/negatives. As in the lecture videos/notes, we'll do a proof by algorithm. Here is the algorithm:

- For $i \in [N]$:
  - If all but E of i's tests come up positive, declare that i is positive.
  - Otherwise, declare that i is negative.

Now we prove that this algorithm works. Suppose that i is indeed positive. Then all of i's tests *should* come up positive, but there might be E false negatives, so all but E tests will come up positive, and we will say that i is positive. Now suppose that i were negative, and Λ is the set of true positives. Then by the disjunctness requirement, there are at least $2E+1$ tests that i is involved in that *should* come up negative. At most E of these can come up positive due to the false positives. So there are still $E+1$ tests that i is involved in that come up negative. Therefore we do not declare i to be positive.

(b) Adapt the Kautz-Singleton argument to show that RS-code-based group testing schemes can handle false positives/negatives. How do the parameters depend on E? (Note: you don't need to change the construction, just the parameters). Your final answer should be of the form "the number of tests T needs to be at least [some function of N, d, and E]."

Solution

Copying the K-S argument, let C be an RS code with dimension k and length $n = q$. Consider the matrix $\Phi \in \{0,1\}^{T \times N}$ where $N = q^k$ items, and $T = q^2$. Thus, we have $k = \log_q(N)$ and $q = \sqrt{T}$.

Let Λ be any set and let i be any other item. The i'th column of Φ can agree with any other in at most $k-1$ places, by the distance of the RS code. Thus, provided that $q \geq dk + 2E + 1$, there are at least $2E+1$ evaluation points of the RS code where codeword i does not agree with any of the codewords in Λ, which translates to there being at least $2E+1$ elements j of $[T]$ so that $\Phi_{j,i} = 1$ and $\Phi_{j,r} = 0$ for all $r \in \Lambda$. (I am omitting some details here, it is exactly the same as the argument in the lecture notes). Thus, if $q \geq d(k-1) + 2E + 1$, our testing matrix is (d, E)-disjunct.

Working out the parameters, we need

$$\sqrt{T} = q \geq d(k-1) + 2E + 1 = d\log_q(N) + 2E + 1$$

or

$$T \geq \left(d(\log_q(N) - 1) + 2E + 1\right)^2.$$

As in class, we have $q \geq d$, so it suffices to take

$$T \geq \left(d(\log_d(N) - 1) + 2E + 1\right)^2.$$

Notice that if E is small compared to $d\log_d(N)$, this doesn't asymptotically affect the answer that we got before with no false positives/negatives. However, if $E \gg d\log_d(N)$, then the $T \geq E^2$ term starts to dominate.

3. **(Bonus – if you finish early, here's something else to work on!)** Can you come up with a way to set parameters in the Kautz-Singleton construction to get good results when, say, $d = N/100$? (Notice that the bound of $d^2 \log N$ isn't great in this parameter regime...) What's the best group testing scheme you can come up with in this setting? (Don't worry about false postives/negatives). What's a natural lower bound on the number of tests you would need?

> **Solution**
>
> This one's a bit open-ended. The KS construction doesn't work well. A natural lower bound is $\log\binom{N}{d} \approx \log((eN/d)^d) = \frac{N}{100} \cdot \log(100 \cdot e)$ bits. I'm actually not sure what the best construction is here!

4. **(Bonus – if you finish early, here's yet another thing to work on!)** Say that a group testing matrix $\Phi \in \{0,1\}^{t \times N}$ is "d-good" if it can identify up to d defective items. More precisely, for $d < N$, $\Phi \in \{0,1\}^{t \times N}$ is d-good iff the map from sets $T \subset [N]$ with $|T| \le d$ to outcomes in $\{0,1\}^t$ given by

$$T \mapsto \left(\bigvee_{i \in T} \Phi_{1,i}, \bigvee_{i \in T} \Phi_{2,i}, \ldots, \bigvee_{i \in T} \Phi_{t,i} \right)$$

is injective.

In class we proved that if $\Phi \in \{0,1\}^{t \times N}$ is d-*disjunct,* then it is d-good.

(a) Show that for $d = 2$, there are matrices that are d-good but not d-disjunct. (It's okay if you show this by giving a somewhat silly example).

(b) Show that any d-good matrix is $(d-1)$-disjunct.

(c) Can you come up with a family of d-good matrices that are not d-disjunct for general d (and which *isn't* a somewhat silly example)?

> **Solution**
>
> (a) Consider
>
> $$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$
>
> This matrix is 2-good, since the possible outcomes are:
>
> $$\begin{aligned} \emptyset &\to (0,0,0,0,0) \\ (1,0,0) &\to (1,0,1,0,0) \\ (0,1,0) &\to (0,1,0,1,1) \\ (0,0,1) &\to (1,1,0,0,1) \\ (1,1,0) &\to (1,1,1,0,0) \\ (1,0,1) &\to (1,1,1,0,1) \\ (0,1,1) &\to (1,1,0,1,1) \end{aligned}$$
>
> and all of these outcomes are different. However, it's not 2-disjunct, since the third column is covered by the union of the first two. This is a bit silly since it's tall and skinny. If you want to make this example less silly, you can do that: if the matrix above is called M, then consider the block matrix
>
> $$\begin{bmatrix} M & 0 \\ 0 & \bar{\Phi} \end{bmatrix}$$
>
> where $\bar{\Phi}$ is a large 2-disjunct matrix. Then you'll get a matrix that is short and fat and still serves as a counter-example.

(b) Suppose that Φ is d good. Let $T \subseteq [N]$ be any set of size at most $d-1$, and let $i \notin T$ be any other index. Then by the definition of good, the outcomes of the tests for T and for $T \cup \{i\}$ are distinct. But this means that there's some index j so that $\bigvee_{\ell \in T} \Phi_{j,\ell} = 0$ and $\bigvee_{\ell \in T \cup \{i\}} \Phi_{j,\ell} = 1$, which means that $\Phi_{j,i} = 1$ and $\Phi_{j,\ell} = 0$ for all $\ell \in T$. Thus, Φ is $(d-1)$-disjunct.