Lecture 15: Perfect Codes & Gilbert-Varshamov Bound

Setting

- Suppose we are given a target distance d
- We are asked to choose a code $\mathcal{C} \subseteq \{0,1\}^n$ with distance d
- Our goal is to maximize $|\mathcal{C}|$

We will see two results:

- We will prove an upper-bound on how large $|\mathcal{C}|$ can be
- We will construct codes that are very large

Definition (Ball)

Let $\mathbb{F}$ be a field of size q. The ball of radius r, represented by $\mathsf{Ball}_q(n, r)$ is the set of all elements in $\mathbb{F}^n$ that have weight $\leqslant r$.

The size of $\mathsf{Ball}_q(n, r)$ is represented by $\mathsf{Vol}_q(n, r)$.

Note that we have

$$\mathsf{Vol}_2(n, r) = \sum_{i=0}^{r} \binom{n}{i}$$

Think: Generalize to arbitrary q.

Definition (Convolution)

Let A and B be two subsets of $\mathbb{F}^n$. By $A + B$ we represent the set $\{a + b \colon a \in A, b \in B\}$.

If $A = \{a\}$, then we write $a + B$ to represent the set $A + B$.

Note that given $x \in \mathbb{F}^n$, the set of all elements in $\mathbb{F}^n$ that are at distance $\leqslant r$ from x is $x + \mathrm{Ball}_q(n, r)$.

Upper Bound

Suppose we have a code $\mathcal{C} \subseteq \{0,1\}^n$ with distance d

Claim

For two distinct codewords $c, c' \in \mathcal{C}$, we have

$$\left(c + \mathrm{Ball}_2(n,r)\right) \cap \left(c' + \mathrm{Ball}_2(n,r)\right) = \emptyset,$$

where $r = \left\lfloor \frac{d-1}{2} \right\rfloor$

- Suppose not
- There exists x such that $d_H(c,x) \leqslant r$ and $d_H(c',x) \leqslant r$
- By triangle inequality, we have $d_H(c,c') \leqslant 2r < d$

- Given this claim, we can conclude that each $c + \mathrm{Ball}_2(n, r)$, where $c \in \mathcal{C}$, is disjoint
- So, we have

$$
\begin{aligned}
\left|\mathcal{C} + \mathrm{Ball}_2(n, r)\right| &= \left|\bigcup_{c \in \mathcal{C}} c + \mathrm{Ball}_2(n, r)\right| \\
&= \sum_{c \in \mathcal{C}} \left|c + \mathrm{Ball}_2(n, r)\right| \\
&= |\mathcal{C}| \cdot \left|\mathrm{Ball}_2(n, r)\right|
\end{aligned}
$$

- Since, $\left|\mathcal{C} + \mathrm{Ball}_2(n, r)\right| \leqslant \left|\{0, 1\}^n\right| = 2^n$, we have the following result

Theorem

Let $\mathcal{C} \subseteq \{0,1\}^n$ *and* $d(\mathcal{C}) = d$*. Then the following holds*

$$|\mathcal{C}| \leqslant \frac{2^n}{\left|\mathrm{Ball}_2(n, r)\right|},$$

where $r = \left\lfloor \frac{d-1}{2} \right\rfloor$*.*

Definition (Perfect Codes)

Codes $\mathcal{C} \subseteq \{0,1\}^n$ with $d(\mathcal{C}) = d$ such that

$$|\mathcal{C}| = \frac{2^n}{\left|\mathrm{Ball}_2(n, r)\right|},$$

where $r = \left\lfloor \frac{d-1}{2} \right\rfloor$, are called Perfect Codes

We state the following theorem without proof. It provides the characterization of all binary linear perfect codes.

Theorem (Tietavainen and Van Lint)

The only binary linear perfect codes are

- *Trivial Codes:* $\{0^n\}$, $\{0,1\}^n$, *and* $\{0^n, 1^n\}$ *for odd* n,
- $[2^r - 1, 2^r - r - 1, 3]_2$ *Hamming Code, and*
- $[23, 12, 7]_2$ *Golay Code.*

Think: Generalize to $\mathcal{C} \subseteq \mathbb{F}^n$.

Suppose we are asked to generate a large code $\mathcal{C} \subseteq \{0,1\}^n$ such that $|\mathcal{C}| = d$. We propose a greedy strategy to generate this code. Consider the following algorithm

1. Let $\mathcal{C} = \emptyset$
2. While $(\{0,1\}^n \setminus \left(\mathcal{C} + \mathrm{Ball}_2(n, d-1)\right) \neq \emptyset)$:
    1. Pick any $c \in \{0,1\}^n \setminus \left(\mathcal{C} + \mathrm{Ball}_2(n, d-1)\right)$
    2. Add c to $\mathcal{C}$
3. Return $\mathcal{C}$

Gilbert-Varshamov Bound II

Theorem (Gilbert-Varshamov Bound)

There exists a code $\mathcal{C}$ with distance d and size $\geqslant \left\lceil \frac{2^n}{\mathrm{Vol}_2(n,d-1)} \right\rceil$

- Our greedy algorithm produces one such code
- The distance is trivially true, because all codewords that are added are at distance $\geqslant d$ from all previous codewords
- If $|\mathcal{C}| < \frac{2^n}{\mathrm{Vol}_2(n,d-1)}$ then $\mathcal{C} + \mathrm{Ball}_2(n, d-1)$ has size $< 2^n$. So, we can choose more codewords

Gilbert-Varshamov Bound III

We can, in fact, choose a binary linear code using a greedy algorithm and achieve the GV-Bound

1. $V = \emptyset$
2. $\mathcal{C}$ be the code spanned by V
3. While $\left(\left(\{0,1\}^n \setminus \mathcal{C} + \mathrm{Ball}_2(n, d-1)\right) \neq \emptyset\right)$:
   1. Pick any v in $\{0,1\}^n \setminus \mathcal{C} + \mathrm{Ball}_2(n, d-1)$
   2. Add v to V
   3. Let $\mathcal{C}$ be the code spanned by V
4. Return $\mathcal{C}$

Prove the following result

Theorem

There exists an $[n,k,d]_2$ binary linear code, where $k \geqslant \left\lceil \lg \frac{2^n}{\mathrm{Vol}_2(n,d-1)} \right\rceil$.

Gilbert-Varshamov Bound IV

In fact, we can randomly create a generator matrix that (roughly) achieves this bound. This has been posed as a homework problem

Generalize all these result to $\mathcal{C} \subseteq \mathbb{F}^n$.