Lecture 15: Perfect Codes & Gilbert-Varshamov Bound
Setting

Suppose we are given a target distance $d$

We are asked to choose a code $C \subseteq \{0,1\}^n$ with distance $d$

Our goal is to maximize $|C|$

We will see two results:

- We will prove an upper-bound on how large $|C|$ 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 $\text{Ball}_q(n, r)$ is the set of all elements in $\mathbb{F}^n$ that have weight $\leq r$.

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

Note that we have

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

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 \{ $a + b$: $a \in A$, $b \in B$ \}.

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 $\leq r$ from $x$ is $x + \text{Ball}_q(n, r)$. 
Suppose we have a code \( C \subseteq \{0, 1\}^n \) with distance \( d \)

**Claim**

*For two distinct codewords \( c, c' \in C \), we have*

\[
(c + \text{Ball}_2(n, r)) \cap (c' + \text{Ball}_2(n, r)) = \emptyset,
\]

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

- Suppose not
- There exists \( x \) such that \( d_H(c, x) \leq r \) and \( d_H(c', x) \leq r \)
- By triangle inequality, we have \( d_H(c, c') \leq 2r < d \)
Given this claim, we can conclude that each \( c + \text{Ball}_2(n, r) \), where \( c \in C \), is disjoint.

So, we have

\[
|C + \text{Ball}_2(n, r)| = \left| \bigcup_{c \in C} c + \text{Ball}_2(n, r) \right| \\
= \sum_{c \in C} |c + \text{Ball}_2(n, r)| \\
= |C| \cdot |\text{Ball}_2(n, r)|
\]

Since, \( |C + \text{Ball}_2(n, r)| \leq |\{0, 1\}^n| = 2^n \), we have the following result.
Theorem

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

$$|C| \leq \frac{2^n}{|\text{Ball}_2(n, r)|},$$

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

Definition (Perfect Codes)

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

$$|C| = \frac{2^n}{|\text{Ball}_2(n, r)|},$$

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

1. Let $C = \emptyset$
2. While $(\{0, 1\}^n \setminus (C + \text{Ball}_2(n, d - 1)) \neq \emptyset)$:
   1. Pick any $c \in \{0, 1\}^n \setminus (C + \text{Ball}_2(n, d - 1))$
   2. Add $c$ to $C$
3. Return $C$
Theorem (Gilbert-Varshamov Bound)

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

- Our greedy algorithm produces one such code
- The distance is trivially true, because all codewords that are added are at distance $\geq d$ from all previous codewords
- If $|C| < \frac{2^n}{\text{Vol}_2(n,d-1)}$ then $C + \text{Ball}_2(n, d - 1)$ has size $< 2^n$. So, we can choose more codewords
We can, in fact, choose a binary linear code using a greedy algorithm and achieve the GV-Bound.

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

Prove the following result

**Theorem**

There exists an \([n, k, d]_2\) binary linear code, where

\[
k \geq \left\lceil \lg \frac{2^n}{\text{Vol}_2(n, d-1)} \right\rceil.
\]
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 $C \subseteq \mathbb{F}^n$. 