Today

- Shannon's Converse Coding Thm
- What can & can't we do
  - Hamming Bd
  - Gilbert-Varshamov

CSS.318.1
Coding Theory
Lecture 4 (2022-9-7)
Instructor: Prahladh Harsha.

Recap from last time

Shannon's Converse Coding Theorem (for BSC)

$\forall p \in (0, 1/2),\ \varepsilon \in (0, \frac{1}{2} - p)$

$\exists \delta, n_0\ \forall n \geq n_0$

If $k \geq (1 - H(p) + \varepsilon) n$ &

$\forall\ E : \{0,1\}^k \to \{0,1\}^n$

$D : \{0,1\}^n \to \{0,1\}^k \cup \{\perp\}$

$\exists m \in \{0,1\}^k$

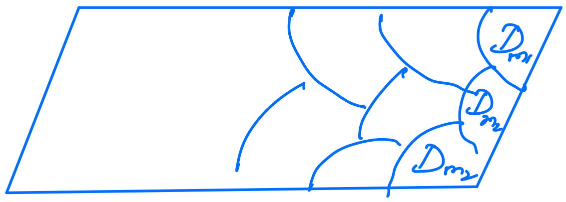
$$\Pr_{e \sim (Ber(p))^n}\left[D(E(m) + e) = m\right] \leq 2^{-\delta n}$$

Proof: Assume (for contradiction) that

$\forall m \in \{0,1\}^k$

$$\Pr_{e}\left[D(E(m) + e) = m\right] \geq 2^{-\delta n}$$

Consider the space



$\{0,1\}^n$ $\quad \forall m \in \{0,1\}^k$

$$D_m = \{ y \in \{0,1\}^n \,/\, D(y) = m \}$$

Obs: D_m's are disjoint

Qn. Do we have lower bd on size of each D_m?

By Assumption $\forall m$, $\quad \Pr_e [E(m)+e \in D_m] \geq 2^{-\delta n}$

ie $\quad \Pr_e [e \in D_m + E(m)] \geq 2^{-\delta n}$

$$S + x = \{ y + x \,/\, y \in S \}$$

$$S_m = \{ y \in \{0,1\}^n \,/\, y \in B(E(m), (p+\epsilon)n) \setminus B(E(m), (p-\epsilon)n) \}$$

$$\Pr_e [E(m) + e \notin S_m] \leq 2^{-C\epsilon^2 n}$$

$$\Pr_e [e + E(m) \in D_m \cap S_m] \geq 2^{-\delta n} - 2^{-C\epsilon^2 n}$$

$\forall m.$

On the other hand.

$$\Pr_e [e + E(m) \in D_m \cap S_m] \leq p_{max} \cdot |D_m \cap S_m|$$

$$p_{max} = \max_{e \in (D_m \cap S_m) + E(m)} \Pr[e]$$

$$\leq \max_{d \in ((p-\varepsilon)n, (p+\varepsilon)n)} p^d (1-p)^{n-d}$$

$$\leq p^{(p-\varepsilon)n} (1-p)^{n-(p-\varepsilon)n} \quad \left[\text{Since } p^d(1-p)^{n-d} = \left(\tfrac{p}{1-p}\right)^d (1-p)^n\right]$$

$\forall m$

$$|D_m \cap S_m| \geq \frac{2^{-\delta n} - 2^{c\varepsilon^2 n}}{p^{(p-\varepsilon)n} (1-p)^{n-(p-\varepsilon)n}}$$

$$= \frac{2^{-2\delta n}}{p^{pn}(1-p)^{(1-p)n}} \left(\frac{p}{1-p}\right)^{\varepsilon n} \quad (\text{if} \cdot$$

$$2^n \geq \sum_m |D_m|$$

$$\geq \sum_m |D_m \cap S_m|$$

$$\geq \sum_m \frac{2^{-2\delta n}}{} \cdot 2^{-H(p)n} \left(\frac{p}{1-p}\right)^{\varepsilon n}$$

$$= 2^{k - H(p)n - n\left(2\delta + \varepsilon \log\left(\frac{1}{p} - 1\right)\right)}$$

if $k > n\left(1 - H(p) - (2\delta + \varepsilon \log(\frac{1}{p} - 1)\right)$

contradiction.

$\boxtimes$

Concluding,

Shannon vs Hamming.

modelling	probabilistic	worst-case # of errors
Code	In terms of E & D (probabilistic)	Explicit code C (decoding explicit)

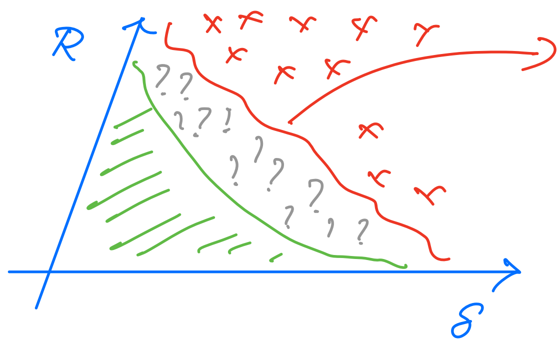
—

Return to Hamming model

Rate vs Distance tradeoff (R vs δ)

$\mathcal{C}$ - $(n, k, d)_q$

$R = k/n$, $\delta = d/n$, Fix q (typically 2)



Impossibility Results

Hamming Bound: $|\mathcal{C}| \leq \dfrac{q^n}{Vol_q\left(n, \left\lfloor \frac{d-1}{2} \right\rfloor\right)}$

Rewriting in terms of R & δ.

$$q^{Rn} \leq \frac{q^n}{q^{\left(h_q\left(\frac{\delta}{2}\right) - o(1)\right)n}} \quad \text{ie,} \quad R \leq 1 - h_q\left(\tfrac{\delta}{2}\right) + o(1)$$

Recall

Claim: $\forall p \in (0, \tfrac{1}{2})$, $2^{(H(p)-o(1))n} \le Vol_2(n, pn) \le 2^{H(p)n}$

Similarly for large q.

$\forall p \in (0, 1-\frac{1}{q})$, $q^{(h_q(p)-o(1))n} \le Vol_q(n, pn) \le q^{h_q(p)n}$

where

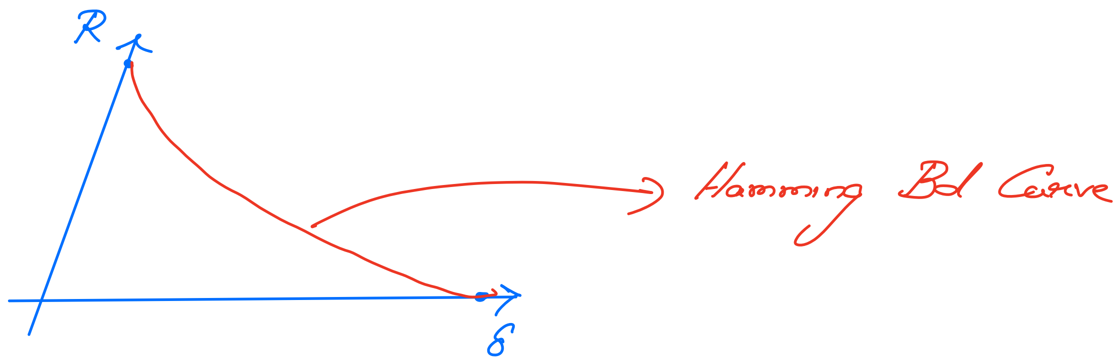
$$h_q(p) = p\log_q \frac{1}{p} + (1-p)\log_q \frac{1}{1-p} + p\log_q(q-1)$$

$$Vol_q(n,t) = \sum_{j=0}^{t} \binom{n}{j}(q-1)^j$$

Hamming Bd (in asymptotic sense)

$$R + h_q\left(\frac{\delta}{2}\right) \le 1 - o(1)$$

$\underline{\underline{q=2}}$



On the other hand, given a δ, how do we construct a code w/ a large R as possible

- Probabilistically (not true exactly but can be modified, pset)
- Greedily (in class)

Greedy Construction: Given n, $d = \delta n$

$$
\left.
\begin{array}{l}
\mathcal{C} \leftarrow \phi, \quad S \leftarrow \{0,1\}^n \\
\text{While } S \neq \phi \\
\quad \left\{ \begin{array}{l} \text{Pick } x \in S \text{ arbitrarily} \\ \mathcal{C} \leftarrow \mathcal{C} \cup \{x\} \\ S \leftarrow S \setminus B(x, d-1) \end{array} \right. \\
\text{Output } \mathcal{C}
\end{array}
\right\}
\begin{array}{l} \text{By design} \\ \Delta(\mathcal{C}) \geq d. \end{array}
$$

What is $|\mathcal{C}|$?

At the end of process

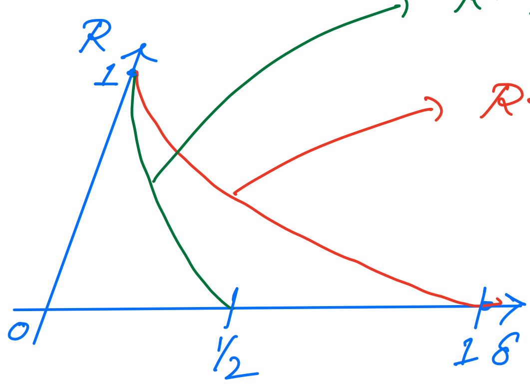
$$B(\mathcal{C}, d-1) = \{0,1\}^n$$

ie,
$$
\begin{aligned}
2^n &= |B(\mathcal{C}, d-1)| \\
&\leq |\mathcal{C}| \cdot Vol(n, d-1)
\end{aligned}
$$

$$|\mathcal{C}| \geq \frac{2^n}{Vol_2(n, d-1)}$$

Rewriting in terms of R & δ.

$$2^{Rn} \geq \frac{2^n}{2^{h(\delta)n}} \quad \text{ie,} \quad \boxed{R + h_2(\delta) \geq 1}$$



$R + H_2(\delta) = 1$ (GV bound).

$R + H_2(\frac{\delta}{2}) = 1$ (Hamming)

[Varshamov]
Gilbert Bd is also achieved by linear codes

Varshamov: linear code

- probabilistic (in class)
- greedy (in pset)

Prob. Const of linear codes:

Pick $G \in \{0,1\}^{n \times k}$ matrix randomly
(ie each entry are picked iid from Ber(1/2))

$$\mathcal{C} = \{ Gx \mid x \in \{0,1\}^k \}$$

- $\dim(\mathcal{C}) = k$
- distance of $\mathcal{C} \geq d$

} for a particular choice of k.

Fix $x \in \{0,1\}^k \setminus \bar{0}$ (non-zero x)

$Gx \sim$ uniformly on $\{0,1\}^n$ (for non-zero x)

$$\Pr_G[wt(Gx) < d] = \frac{Vol(n, d-1)}{2^n}$$

$$\Pr[\exists \text{ non-zero } x, \; wt(Gx) < d] \leq 2^k \cdot \frac{Vol(n, d-1)}{2^n}$$

If $k < n - h(p)n$, < 1

ie, with high probability a random G (s.t. $k < n - h(p)n$)

is the generator matrix of a $[n, k, d]$-code.