Rohan Garg

Error Correcting Codes

Lecture 11 - CMU Toolkit

- Setting / Preliminaries of Error Correcting Codes
- Linear Error-Correcting Codes
- Hamming and Hadamard Codes
- Reed-Solomon Code

Basic Definitions:

Def: An Error-Correcting code is an injective map from k-length strings to n-length strings:

$$Enc: \Sigma^k \longrightarrow \Sigma^n$$

where Σ is the alphabet. We will generally take $\Sigma = \{0,1\}$

- $q = |\Sigma|$. $q = 2 \Rightarrow$ binary

- Message: k is message dimension

  elements in Σ^k are messages

- Block length: msg $\longrightarrow$ n-bit string

- Code: # Codes $= q^k$

- Rate $= \frac{k}{n}$. $\frac{|msg|}{|block|}$ Ideally, this should be close to 1.

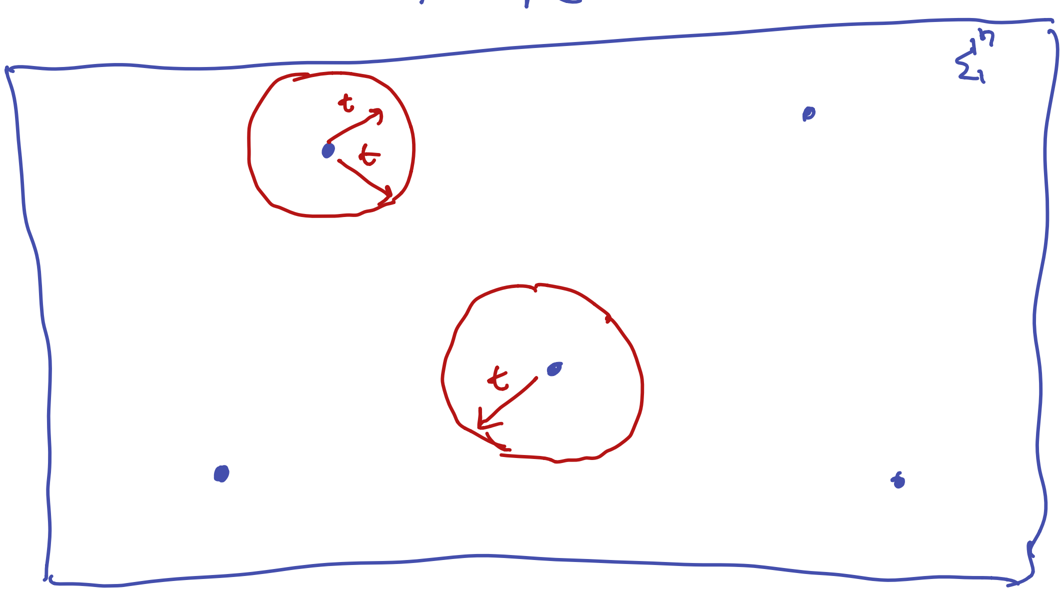
$$\text{MSG} \longrightarrow \text{Enc}(\text{MSG}) \xrightarrow{\leq t \text{ errors}} Z \xrightarrow{\text{recover?}} \text{MSG}$$

$$(\Sigma^k) \qquad\qquad (\Sigma^n) \qquad\qquad (\Sigma^n) \qquad\qquad (\Sigma^k)$$

Hamming Distance:

Def: (Hamming Dist.): Number of positions at which two strings differ.

$$\Delta(x,y) = \left| \{ i : x_i \neq y_i \} \right|$$



d the minimum distance between any 2 vertices.

$$d = \min_{y \neq y'} \left\{ \Delta(y, y') \right\}$$

Fact: Unique decoding (for each z the receiver gets, there is a unique x she can recover) is possible iff $t \leq \left\lfloor \frac{d-1}{2} \right\rfloor$.

LINEAR CODE

A linear code of length n and rank k is a linear subspace C with dimension k of the vector space $\mathbb{F}_q^n$ where $\mathbb{F}_q$ is the finite field of q elements.

Def: (Linear Code). $\text{Enc}: \mathbb{F}_q^k \to \mathbb{F}_q^n$

$$x \to Gx$$

where x is a vector and G is a matrix.

G is called the "Generator matrix".

→ full-rank $n \times k$ matrix

$C = \text{Im}(G)$ = image of G which spans all linear combinations of rows.

Notation:

$[n, k, d]_q$ linear code

(a, b, c) not necessarily linear

n = length of codeword

K = length of message

d = min. distance

$q = |\Sigma|$.

$z = Gx$

Let $C^{\perp} = \{ w \in F_q^n : w^T z = 0, \forall z \in C \}$

→ every vector in $C^{\perp}$ is orthogonal to every codeword (vector in C).

$C^{\perp} \rightarrow [n, n-k]_q$ code.

$\mathrm{Enc}^{\perp} : \mathbb{F}_q^{n-k} \longrightarrow \mathbb{F}_q^{n}$ maps w to $H^T w$

H is a $(n-k) \times n$ matrix:

Parity check matrix.

$$\begin{bmatrix} \longleftarrow & \longrightarrow \\ \longleftarrow & \longrightarrow \\ & H \\ \longleftarrow & \longrightarrow \end{bmatrix}$$

$C^{\perp}$ is rowspan of H:

$$z \in C \iff Hz = \vec{0}$$

<u>Def:</u> Hamming Weight $wt(w) = \Delta(w, \vec{0})$.

<u>Fact:</u> $d(C)$ is the least Hamming Wt. of a non-zero codeword.

$$\Rightarrow \Delta(y, y') = wt(y - y')$$

<u>Fact:</u> $d(C)$ = min number of columns in H which are linearly dependent.

<u>Proof:</u>

$$d(C) = \min\{wt(z) : z \in C, z \neq 0\}$$

$$= \min\{wt(z) : Hz = 0, z \neq 0\}$$

Hamming Code:

$q = 2$, binary set up:

$$H = \underbrace{\left.\begin{bmatrix} 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \cdots & 1 \\ \vdots & \vdots & \vdots & \cdots & 1 \\ 0 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 1 & \cdots & 1 \end{bmatrix}\right\} r}_{2^r - 1}$$

H is $r \times 2^r - 1$ matrix and the columns span all possible binary strings of length r. [except zero column]

H is full-rank because it has the identity matrix.

distance for $H = 3$.

$$\text{Ham} \rightarrow [2^r - 1,\ 2^r - 1 - r,\ 3]_2$$

let $n = 2^r$

$$[\underset{\substack{\uparrow \\ \text{block}}}{n},\ \underset{\substack{\uparrow \\ \text{msg}}}{n - \log(n+1)},\ 3]$$

Rate: :)

distance: :($\quad 3 \rightarrow \left\lfloor \frac{d-1}{2} \right\rfloor$ errors

$\rightarrow$ handle 1 error.

z n-length string.

if: $Hz = 0 \longrightarrow$ msg was not modified

else:

$$e_i = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ 0 \end{bmatrix} \leftarrow i^{th} \text{ coordinate}$$

For some i,

$$z = \text{msg} + e_i$$

$$Hz = \underbrace{H \cdot msg}_{\substack{\parallel \\ 0}} + He_i = He_i$$

$$H = \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 & \cdots & \cdots \end{matrix} \\ \begin{bmatrix} 0 & 0 & & & & 1 \\ 0 & 0 & & & & 1 \\ 0 & 0 & \cdots & & & 1 \\ 0 & 1 & & & & 1 \\ \vdots & \vdots & & & & \vdots \\ 1 & 0 & & & & 1 \end{bmatrix} \end{array} \begin{bmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ 0 \end{bmatrix} \rightarrow$$

Perfect Code: A perfect code may be interpreted as one in which the "balls" of radius t exactly fill out the space.

→ Good rate, Bad distance

(# errors tolerated)

Hadamard Code:

$$H^T = \begin{bmatrix} 0\,0 \cdots\cdots 0\,1 \\ 0\,0 \cdots\cdots 1\,0 \\ 0\,0 \cdots\cdots 1\,1 \\ \vdots \\ 1\,1\,1\,1\,1\,1\,1\,1\,1\,1 \end{bmatrix}$$

→ Add in zero's row!

$$G = \begin{bmatrix} 0\,0 \cdots\cdots 0\,0 \\ \\ H^T \\ \\ \end{bmatrix}$$

Generator matrix for Hadamard Code.

Def: Hadamard Code. Hadamard encoding of x is defined as the sequence of all inner products with x:

$$x \longrightarrow (a \cdot x)_{a \in \mathbb{F}_2^r}$$

Given: $x \in \mathbb{F}_2^r$, define r-variate linear polynomial

$$L_x : \mathbb{F}_2^r \to \mathbb{F}_2$$

$$\Rightarrow \quad a \longrightarrow x^T a = \sum_{i=1}^{r} x_i a_i$$

x_i's $\leftarrow$ coefficients

a_i's $\leftarrow$ variables

"Like" : mapping x to the truth table of L_x:

$$(a\cdot x)_{a\in\mathbb{F}_2^r} = \left(L_x(a)\right)_{a\in\mathbb{F}_2^r}$$

<u>Fact:</u> Hadamard Code is a $[\underset{\text{block}}{2^r}, \underset{\text{msg}}{r}, \underset{\text{distance}}{2^{r-1}}]_2$ code.

Let $n = 2^r$

$\rightarrow$ $[n, \log(n), \frac{1}{2}n]_2$ code.

:) : # errors [distance]

:(: rate $\quad \dfrac{\text{msg}}{\text{block}} \quad \dfrac{\log(n)}{n}$

Reed-Solomon Codes (RS):

Super Useful!!

<u>Def: (RS code):</u> For $1 \leq k < n$, $q \geq n$,

select a subset of symbols of cardinality n,

$S \subseteq \mathbb{F}_q$, $|S| = n$.

Enc: $\mathbb{F}_q^k \rightarrow \mathbb{F}_q^n$:

message $m = m_0 m_1 \cdots m_{k-1}$

$$m \rightarrow \left(P_m(a)\right)_{a \in S}$$

$P_m(a) \in \mathbb{F}_q[x] = m_0 + m_1 x + \ldots + m_{k-1} x^{k-1}$

Facts:

- Linear Code:

$$\text{Enc}(m+m') = \text{Enc}(m) + \text{Enc}(m')$$

→ adding coeffecients of polynomial.

- Generator matrix:

each row is $[1, a, a^2, \ldots, a^{k-1}]$ for some $a \in S$

"Vandermonde" matrix.

- min-dist $\geq n-(k-1) = n-k+1$.

Bad property: $q \geq n$.

→ $[n, k, n-k+1]_q$

→ optimal for these parameters.

Theorem: Singleton Bound:

For a $[n,k,d]_q$ code,

$$k \leq n-d+1 .$$ 1964.

Thanks!