Lecture 16: Introduction to Error-correcting Codes

Basic Definitions

Definition (Hamming Distance)

The Hamming distance between two strings $x, y \in \Sigma^n$, denoted by $\Delta(x, y)$, is the number of positions $i \in [n]$ such that $x_i \neq y_i$
Relative Hamming distance between x, y is represented by $\delta(x, y) := \Delta(x, y)/n$.

Definition (Hamming Weight)

The Hamming weight of a strings $x \in \Sigma^n$, denoted by $\mathrm{wt}(x)$, is the number of non-zero symbols in x.

- Note that $\Delta(x, y) = \mathrm{wt}(x - y)$
- Hamming ball of radius r around x is the set $\{y \colon y \in \Sigma^n, \Delta(x, y) \leqslant r\}$

Basic Definitions

Definition (Error-correcting Code)

An error-correcting code C is a subset of Σ^n

- If $|\Sigma| = q$, then the code C is called q-ary code
- The block-size of code C is n
- Encoding map is a mapping of the set of messages $\mathcal{M}$ to C

Rate and Dimension

Definition (Rate of a code)

The rate of a code is defined:

$$R(C) := \frac{\log |C|}{n \log |\Sigma|}$$

- The dimension of a code is defined to be $\log |C| \,/\, \log |\Sigma|$

Distance

Definition (Distance)

The distance of a code C is:

$$\Delta(C) := \min_{\substack{c_1, c_2 \in C \\ c_1 \neq c_2}} \Delta(c_1, c_2)$$

- The relative distance of a code is $\delta(C) = \Delta(C)/n$

Examples

- Repetition code repeats every input bit t times. It has block-size n, dimension n/t and distance t.
- Parity-check code appends the parity of $(n-1)$ bits at the end. It has block-size n, dimension $(n-1)$ and distance 2.
- Hamming code encodes 4 bits (x_1, x_2, x_3, x_4) as $(x_1, x_2, x_3, x_4, a, b, c)$, where $a = x_2 + x_3 + x_4$, $b = x_1 + x_3 + x_4$ and $c = x_1 + x_2 + x_4$. It has block-size 7, dimension 4 and distance 3.

Lemma

The following statements are equivalent:

- C *has minimum distance* $2t + 1$
- C *can detect* $2t$ *symbol erasures*
- C *can correct* $2t$ *symbol erasures*
- C *can correct* t *symbol errors*

Linear Codes

Definition (Linear Code)

If Σ is a field and $C \subseteq \Sigma^n$ is a subspace of Σ^n then C is a linear code

- If C has dimension k, then there exists codewords $\{c_1, \ldots, c_k\} \subseteq C$ such that any codeword $c \in C$ can be written as linear combination of $\{c_1, \ldots, c_k\}$
- Every codeword can be written as $x \cdot G$, where $x = (x_1, \ldots, x_k) \in \Sigma^k$ and $G = \begin{pmatrix} c_1 \\ \vdots \\ c_k \end{pmatrix} \in \Sigma^{k \times n}$
- G is called the generator matrix of C
- The mapping $x \mapsto xG$ is an encoding map
- A q-ary binary linear code with block length n, with dimension k and distance d is represented by $[n, k, d]_q$

Examples

- Parity-check code is an $[n, n-1, 2]_2$ code
- Repetition code is an $[n, n/t, t]_2$ code
- Hamming code is an $[7, 4, 3]_2$ code

Think: Their generator matrix?

Definition (Systematic Form)

If $G \equiv [I|G']$, G is said to be in the systematic form

Parity Check Matrices

Lemma

C is an $[n, k]_q$ code if and only if there exists a matrix $H \in \mathbb{F}_q^{(n-k)\times n}$ of full row rank such that

$$C = \{c \colon c \in \mathbb{F}_q^n, Hc = 0\}$$

- H is called the parity check matrix for C

Lemma

$\Delta(C)$ equals the minimum number of columns of H that are linearly dependent.

Lemma

If $G = [I|G']$ is in systematic form, then $H = [G'^{\intercal}|I]$ is the parity check matrix.

Example

For Hamming code, we have

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

$$H = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

- H has all non-zero binary strings of length 3 as its columns

Correcting One Error with Hamming code

- Let c be the transmitted codeword
- Let e_i be the error introduced
- Received codeword is $\widetilde{c} = c + e_i$
- Note that $H\widetilde{c} = Hc + He_i = H_i$
- So, we can find the position where error has occurred and it can be removed

Definition (Syndrome)

Hy is the syndrome of y

Generalized Hamming code

- Let $H \in \mathbb{F}_q^{r\times(2^r-1)}$ such that the i-th column is the binary representation of i, for $i \in [2^r - 1]$
- Define C using the parity check matrix

Lemma

C is an $[2^r - 1, 2^r - r - 1, 3]$ code