Problem set 7

Please turn in your problem set at the start of class. You can place it up here on the desk.

COSC 290 Discrete Structures

Lecture 10: Proofs (of properties of error-correcting codes)

Prof. Michael Hay
Wednesday, Sep. 20, 2017
Colgate University

Plan for today

1. Error correcting codes
2. Minimum Distance
3. Hamming codes

Error correcting codes

Error correcting codes: an abstract view

A code is a set $\mathcal{C} \subseteq \{0,1\}^n$ where $|\mathcal{C}| = 2^k$.

- Encoding: a bijective function $encode : \{0,1\}^k \to \mathcal{C}$ maps k-bit messages to codewords in $\mathcal{C}$. Both the sender and receiver know this function.
- Error detection: a function $hasError(c')$ returns True if $c' \notin \mathcal{C}$ and False otherwise.
- Error correction: choose $c \in \mathcal{C}$ that is *closest* (in terms of Hamming distance) to c' and then apply the inverse of *encode*.

Example

Example code where $\mathcal{C} := \{100111, 101010, 010110, 010111\}$. Since $|\mathcal{C}| = 2^2$, we can use code to send 2-bit messages.

m	$c \in \mathcal{C}$
00	10 01 11
01	10 10 10
10	01 01 10
11	01 01 11

Note: the rows of this table define one particular *encode* function.

Definition of error correcting

Let $\mathcal{C} \subseteq \{0,1\}^n$ be a code and let $\ell \geq 1$ be any integer.

We say that code $\mathcal{C}$ can correct ℓ errors if, for any codeword $c \in \mathcal{C}$ and for any sequence of up to ℓ errors applied to c, we can correctly identify that c was the original codeword.

Poll: Interpreting the definition

Suppose that $\mathcal{C}$ can correct correct ℓ errors, meaning that for any codeword $c \in \mathcal{C}$ and for any sequence of up to ℓ errors applied to c, we can correctly identify that c was the original codeword.

Which of the following facts are implied? (You can assume $\mathcal{C}$ is not empty.)

A) For any $c \in \mathcal{C}$, there exists some $c' \in \{0,1\}^n$ where $\Delta(c, c') \leq \ell$ and the receiver, given only c', can correctly identify that c was the original codeword.
B) For any $c \in \mathcal{C}$, for any $c' \in \{0,1\}^n$ if $\Delta(c, c') \leq \ell$, then the receiver, given only c', can correctly identify that c was the original codeword.
C) For any $c \in \mathcal{C}$, there exists some $c' \in \{0,1\}^n$ where $\Delta(c, c') > \ell$ and the receiver, given only c', cannot correctly identify that c was the original codeword.
D) For any $c \in \mathcal{C}$, for any $c' \in \{0,1\}^n$ if $\Delta(c, c') > \ell$, then the receiver, given only c', cannot correctly identify that c was the original codeword.
E) None of above/More than one of above.

Problem set 7

A size 3 repetition code with 4 bit messages can correct ℓ errors for $\ell = 1$.

- This means that for any codeword c, if you introduce at most 1 error, the original codeword c can be still recovered through error correction.
- This does *not* mean that if you introduce 2 or more errors, there is no hope of error correction. Some codewords may be able to tolerate more than one 1 error!

What is the *largest* number of errors that can possibly be corrected successfully by this code?

Can we generalize the result to a size m-repetition code with k-bit messages?

Minimum Distance

Minimum Distance

The minimum distance of code $\mathcal{C}$ is the smallest Hamming distance between two distinct codewords in $\mathcal{C}$.

$$\min \{ \Delta(x, y) : x, y \in \mathcal{C} \text{ and } x \neq y \}$$

m	$c \in \mathcal{C}$
00	10 01 11
01	10 10 10
10	01 01 10
11	01 01 11

Poll: minimum distance

Consider this code $\mathcal{C}$?

m	$c \in \mathcal{C}$
00	10 01 11
01	10 10 10
10	01 01 10
11	01 01 11

What is its minimum distance?

A) 0

B) 1

C) 2

D) 3

The minimum distance of code $\mathcal{C}$ is the smallest Hamming distance between two distinct codewords in $\mathcal{C}$.

$$\min \{ \Delta(x, y) : x, y \in \mathcal{C} \text{ and } x \neq y \}$$

Theorem: minimum distance and detecting/correcting errors

If the minimum distance of a code $\mathcal{C}$ is $2t+1$, then $\mathcal{C}$ can detect $2t$ errors and correct t errors.

Proofs on board

Hamming codes

Hamming code

The Hamming code for a 4-bit message $\langle a, b, c, d\rangle$ is the original message, followed by three parity bits:

$$\langle a, b, c, d, (b \oplus c \oplus d), (a \oplus c \oplus d), (a \oplus b \oplus d)\rangle$$

Example: if message is $\langle 1, 0, 1, 1\rangle$, the codeword is

$$\begin{aligned} &\langle 1, 0, 1, 1, (0 \oplus 1 \oplus 1), (1 \oplus 1 \oplus 1), (1 \oplus 0 \oplus 1)\rangle \\ = &\langle 1, 0, 1, 1, 0, 1, 0\rangle \end{aligned}$$

Hamming code properties

1. Every message bit appears in *at least two* parity bits. Why significant?
   If message bit gets corrupted, at least two parity bits will be off.
   If parity bit gets corrupted, only it will look wrong.
2. No two message bits appear in precisely the *same set* of parity bits. Why significant?
   By looking at which parity bits appear off, you can pinpoint source of error.

Poll: Hamming code error correction

Reminder: The Hamming code for a 4-bit message $\langle a, b, c, d \rangle$ is

$$\langle a, b, c, d, (b \oplus c \oplus d), (a \oplus c \oplus d), (a \oplus b \oplus d) \rangle$$

Question: You receive the following (possibly corrupted) Hamming codeword. Assume at most one error has occurred, find the original message.

$$\langle 1, 0, 1, 1, 1, 1, 1 \rangle$$

Hint: assume the message is *uncorrupted*, figure out what the parity bits *should* be under that assumption, and then if the parity bits don't match the message, try to *pinpoint the error*.

A) The message is... $\langle 0, 1, 1, 1 \rangle$
B) The message is... $\langle 1, 1, 1, 1 \rangle$
C) The message is... $\langle 0, 0, 1, 0 \rangle$
D) The message is uncorrupted, the error is in a parity bit.
E) There is no corruption ($\langle 1, 0, 1, 1, 1, 1, 1 \rangle$ is a valid codeword).

Problem set 7

The second problem on the homework was easily misinterpreted. Two reasonable interpretations:

- Define 11-bit Hamming codeword for 7-bit message with 4 parity bits.
- Define 15-bit Hamming codeword for 11-bit message with 4 parity bits.