Poll: from the reading

Which of the following is the most accurate statement about Hamming codes?

A) in Hamming codes, some but not all message bits are repeated

B) (4,7)-Hamming codes have a lower rate than $\ell = 3$ repetition codes.

C) a (4,7)-Hamming codes can correct at most one error.

D) Hamming codes are the codewords with minimum Hamming distance

COSC 290 Discrete Structures

Lecture 12: Hamming codes

Prof. Michael Hay

Monday, Feb. 19, 2018

Colgate University

Plan for today

1. Minimum distance
2. Hamming codes
3. Minimum distance of Hamming codes

Minimum distance

Recall: error correcting codes

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.
- Sender writes message m, encodes it into codeword c and sends it.
- Receiver receives n-bit message c' and attempts to detect/correct any errors.

Recall: error correcting codes

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.
- Sender writes message m, encodes it into codeword c and sends it.
- Receiver receives n-bit message c' and attempts to detect/correct any errors.
- Error detection: the receiver will conclude an error has occurred if and only if $c' \notin \mathcal{C}$.

Recall: error correcting codes

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.
- Sender writes message m, encodes it into codeword c and sends it.
- Receiver receives n-bit message c' and attempts to detect/correct any errors.
- Error detection: the receiver will conclude an error has occurred if and only if $c' \notin \mathcal{C}$.
- Error correction: receiver chooses $c'' \in \mathcal{C}$ that is *closest* to c'.

Recall: error correcting codes

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.
- Sender writes message m, encodes it into codeword c and sends it.
- Receiver receives n-bit message c' and attempts to detect/correct any errors.
- Error detection: the receiver will conclude an error has occurred if and only if $c' \notin \mathcal{C}$.
- Error correction: receiver chooses $c'' \in \mathcal{C}$ that is *closest* to c'.
- After correction, receiver decodes c'' by applying inverse of *encode*.

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 $\mathcal{C}$.

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

Theorems: 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.

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

Proofs on board, but first a question

Poll: minimum distance

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

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

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

Consider an $\ell = 5$ repetition code applied to $k = 3$-bit messages. How many errors can such a code detect?

A) 0

B) 1

C) 2

D) 3

E) 4 or more

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

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

Poll: what is Hamming codeword for this message?

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

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

Suppose message is $\langle 0, 1, 1, 0 \rangle$, what is the Hamming codeword?

A) $\langle 0, 1, 1, 0, 0, 0, 0 \rangle$

B) $\langle 0, 1, 1, 0, 0, 0, 1 \rangle$

C) $\langle 0, 1, 1, 0, 0, 1, 0 \rangle$

D) $\langle 0, 1, 1, 0, 0, 1, 1 \rangle$

E) None of the above

Hamming code properties

1. Every message bit appears in *at least two* parity bits. Why significant?

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.

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?

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.

Second criterion

No two message bits appear in precisely the *same set* of parity bits.

This is not the same thing as saying no message bit can appear in a parity bit set more than once.

$$\langle a, b, c, d, \underbrace{(b \oplus c \oplus d)}_{\text{parity bit } b_1}, \underbrace{(a \oplus c \oplus d)}_{\text{parity bit } b_2}, \underbrace{(a \oplus b \oplus d)}_{\text{parity bit } b_3}, \rangle$$

Second criterion

No two message bits appear in precisely the *same set* of parity bits.

This is not the same thing as saying no message bit can appear in a parity bit set more than once.

$$\langle a, b, c, d, \underbrace{(b \oplus c \oplus d)}_{\text{parity bit } b_1}, \underbrace{(a \oplus c \oplus d)}_{\text{parity bit } b_2}, \underbrace{(a \oplus b \oplus d)}_{\text{parity bit } b_3}, \rangle$$

message bit	parity bit set
a	$\{2, 3\}$
b	$\{1, 3\}$
c	$\{1, 2\}$
d	$\{1, 2, 3\}$

The criterion no two message bits can have the same set. But each message can (and does) appear in more than parity bit.

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, what is 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).

Minimum distance of Hamming codes

Hamming code

Claim: Hamming code has a minimum distance of 3.

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

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

Poll: What to show?

We want to support claim that the Hamming code $\mathcal{C}$ has a minimum distance of at least 3. Which of the following statements provides sufficient support for our claim? Note: we use *encode* to mean the function that generates a Hamming codeword c for message m.

A) There exists two codewords $c \in \mathcal{C}$ and $c' \in \mathcal{C}$ such that $c \neq c'$ and $\Delta(c, c') = 3$.

B) There exists two messages $m \in \{0,1\}^k$ and $m' \in \{0,1\}^k$, such that $m \neq m'$ and $\Delta(encode(m), encode(m')) = 3$.

C) For any two codewords $c \in \mathcal{C}$ and $c' \in \mathcal{C}$, if $c \neq c'$, then $\Delta(c, c') > 3$.

D) For any two messages $m \in \{0,1\}^k$ and $m' \in \{0,1\}^k$, if $m \neq m'$, then then $\Delta(encode(m), encode(m')) \geq 3$.

E) None of above / More than one

Claim

The claim we want to prove is

$$\forall m \in \{0,1\}^k : \forall m' \in \{0,1\}^k :$$
$$\Delta(m, m') > 0 \implies \Delta(encode(m), encode(m')) \geq 3$$

(where *encode* is the function that generates a Hamming codeword c for message m).

Claim, with slight notation change

The claim we want to prove is

$$\forall m \in \{0,1\}^k : \forall m' \in \{0,1\}^k : \\ \Delta(m, m') > 0 \implies \Delta(c, c') \geq 3$$

(where we use c to denote the Hamming codeword for m and c' to denote Hamming code word for m').

Proof by cases

$$\forall \langle m, m' \rangle \in \left(\{0,1\}^k \times \{0,1\}^k \right) : \Delta(m, m') > 0 \implies \Delta(c, c') \geq 3$$

Cases:

- $\langle m, m' \rangle$ pairs such that $\Delta(m, m') \geq 3$
- $\langle m, m' \rangle$ pairs such that $\Delta(m, m') = 2$
- $\langle m, m' \rangle$ pairs such that $\Delta(m, m') = 1$
- $\langle m, m' \rangle$ pairs such that $\Delta(m, m') = 0$ (trivial case)

Important: when proving by cases, make sure your cases cover all possibilities!

For each case, we need to show $\Delta(m, m') > 0 \implies \Delta(c, c') \geq 3$.

Shown on board

Poll: What to show for second case?

We are considering the case where $\Delta(m, m') = 2$. Recall that we are using $c := \mathit{encode}(m)$ and $c' := \mathit{encode}(m')$. What do we want to show? Choose the best answer.

A) $\Delta(c, c') = 2$

B) $\Delta(c, c') > 3$

C) That c and c' differ in at least one parity bit.

D) That c and c' differ in at least two parity bits.

E) None / More than one