Error-Correcting Codes

Based on a handout by Yingkun Li
Revised by Mark on January 16, 2023

Part 1: Error Detection

An ISBN[1] is a unique numeric book identifier. It comes in two forms: ISBN-10 and ISBN-13. Naturally, ISBN-10s have ten digits, and ISBN-13s have thirteen. The final digit in both versions is a *check digit*.

Say we have a sequence of nine digits, forming a partial ISBN-10: $n_1 n_2 ... n_9$.
The final digit, n_{10}, is calculated as follows:

$$\left(\sum_{i=1}^{9} i \times n_i \right) \bmod 11$$

If n_{10} is equal to 10, it is written as `X`.

Problem 1:
Which of the following could be valid ISBNs?

- `0-134-54896-2`
- `0-307-29206-3`
- `0-316-00395-6`

[1] International Standard Book Number

Problem 2:
Show that the following sum is divisible by 11 iff $n_1 n_2 ... n_{10}$ is a valid ISBN-10.

$$\sum_{i=1}^{10} (11 - i) n_i$$

Problem 3:
Take a valid ISBN-10 and change one digit. Is it possible that you get another valid ISBN-10? Provide an example or a proof.

Problem 4:
Take a valid ISBN-10 and swap two adjacent digits. When will the result be a valid ISBN-10? This is called a *transposition error*.

Problem 5:
ISBN-13 error checking is slightly different. Given a partial ISBN-13 $n_1 n_2 n_3 ... n_{12}$, the final digit is given by

$$n_{13} = \left[\sum_{i=1}^{12} n_i \times (2 + (-1)^i) \right] \bmod 10$$

What is the last digit of the following ISBN-13?
`978-0-380-97726-?`

Problem 6:
Take a valid ISBN-13 and change one digit. Is it possible that you get another valid ISBN-13? Provide an example or a proof.

Problem 7:
Take a valid ISBN-13 and swap two adjacent digits. When will the result be a valid ISBN-13?
Hint: The answer here is more interesting than it was last time.

Problem 8:
`978-0-08-2066-46-6` was a valid ISBN until I changed a single digit.
Can you tell me which digit I changed?

Part 2: Error Correction

Error detection is helpful, but we'd also like to fix errors when we find them. One example of such a system is the QR code, which remains readable even if a significant amount of it is removed. QR codes with icons inside aren't special–they're just missing their central elements. The error-correcting codes in the QR specification allow us to recover the lost data.

Definition 1: Repeating codes
The simplest possible error-correcting code is a "repeating code". It works just as you'd expect: Instead of sending data once, it sends multiple copies. If a few bits are damaged, they can be both detected and repaired.
For example, consider the following three-repeat code encoding the binary string 101:

$$111\ 000\ 111$$

If we flip any one bit, we can easily find and fix the error.

Definition 2: Code Efficiency
The efficiency of an error-correcting code is calculated as follows:

$$\frac{\text{number of data bits}}{\text{total bits sent}}$$

For example, the efficiency of the three-repeat code above is $\frac{3}{9} = \frac{1}{3} \approx 0.33$

Problem 9:
What is the efficiency of an k-repeat code?

Problem 10:
How many repeated digits do you need to...
- detect a transposition error?
- correct a transposition error?

Definition 3: Hamming's Square Code
A more effective coding scheme comes in the form of Hamming's square code. Take a four-bit message and arrange it in a 2×2 square.
Compute the pairity of each row and write it at the right.
Compute the pairity of each column and write it at the bottom.
Finally, compute the pairity of the entire message write it in the lower right corner.

Reading the result row by row to get the encoded message.
For example, the message 1011 generates the sequence 101110011:

$$1011 \longrightarrow \begin{array}{cc|} 1 & 0 \\ 1 & 1 \\ \hline \end{array} \longrightarrow \begin{array}{cc|c} 1 & 0 & 1 \\ 1 & 1 & 0 \\ \hline 0 & 1 & \end{array} \longrightarrow \begin{array}{cc|c} 1 & 0 & 1 \\ 1 & 1 & 0 \\ \hline 0 & 1 & 1 \end{array} \longrightarrow 101110011$$

Problem 11:
The following message are encoded using the method above. Find and correct any single-digit or transposition errors.

1. `110 110 011`
2. `100 101 011`
3. `001 010 110`

Problem 12:
What is the efficiency of this coding scheme?

Problem 13:
Can we correct a single-digit error in the encoded message?
Can we correct a transposition error in the encoded message?

Problem 14:
Let's generalize this coding scheme to a non-square table:
Given a message of length ab, construct a rectangle with dimensions $a \times b$ as described above.

- What is the efficiency of a $a \times b$ rectangle code?
- Can the $a \times b$ rectangle code detect and fix single-bit errors?
- Can the $a \times b$ rectangle code detect and fix two-bit errors?

Part 3: Hamming Distance

Definition 4:
The *Hamming distance* between two strings $x = x_1 x_2 ... x_n$ and $y = y_1 y_2 ... y_n$ is the number of positions at which the digits of x and y are different.

Problem 15:
Compute the Hamming distance between `1010` and `0001`.

Problem 16:
Read $d_H(x, y)$ as "the hamming distance between x and y."
Prove the following statements:

1. $d_H(x, y) \geq 0$ with equality if and only if $x = y$,
2. $d_H(x, y) = d_H(y, x)$,
3. $d_H(x, z) \leq d_H(x, y) + d_H(y, z)$.

Problem 17:
Say we encode and send a message with the 3-repeat code. A few bits are damaged in transit. When the transmission is decoded, a different message is read.

What is the minimum possible hamming distance between the undamaged encoded message and the damaged encoded message?

Problem 18:
Say we encode and send a message with Hamming's square code. A few bits are damaged in transit. When the transmission is decoded, no uncorrectable errors are detected and a different message is read.

What is the minimum possible hamming distance between the undamaged encoded message and the damaged encoded message?

Part 4: Hat Puzzles: The Revenge

Problem 19:
Three people are sitting in a circle. A black or a white hat will be placed on each person's head, with equal probability. Each person can see everyone's hat color except their own.

The participants are asked to simultaneously write down "Black", "White", or "Pass".
They win if at least one person guesses their hat correctly.
They lose if anyone guesses incorrectly, or if everyone passes.

How can they maximize their chance of winning?

Problem 20:
Consider the same game with $2^n - 1$ people. How can they achieve a win rate of $\frac{2^n-1}{2^n}$?