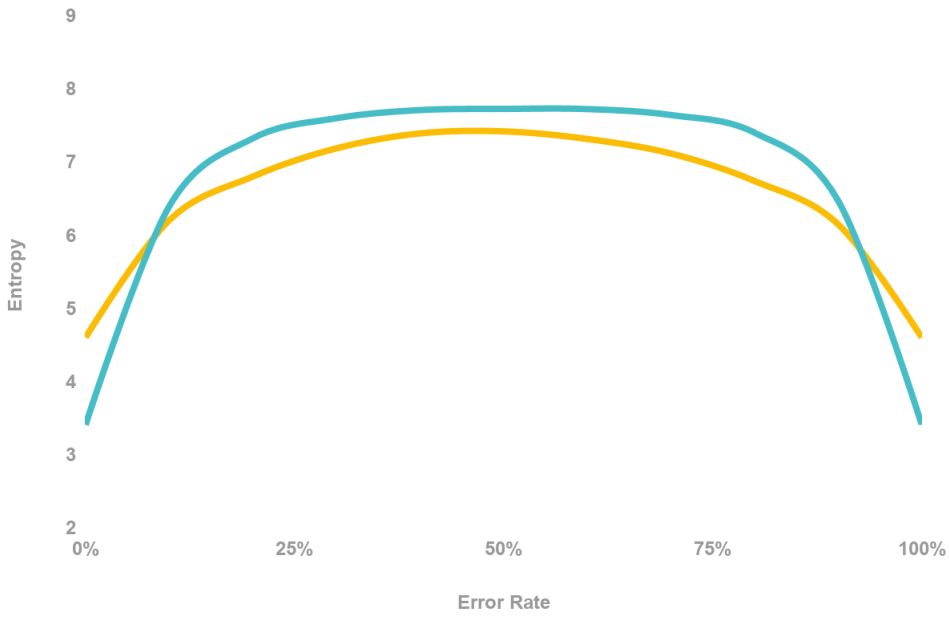
Writeup
Assignment 5

Recap

For assignment 5 I created two programs: one for encoding a file, and another for decoding the file. I was also provided with a program that adds random errors to a given file.

Error Rate and Entropy

What happens to the *entropy* of a file when the *error rate* is increased? When starting from an error rate of 0%, gradually increasing the error rate will rapidly increase the amount of entropy, as expected. For error rates of 10% and above, the file is typically unreadable. However, for extremely high error rates, the *entropy* strangely begins to *decrease*. Moreover, *the entropy at 100% error rate is no higher than the entropy at 0% error rate*. What could explain this? Well, recall that Hamming codes only work because we know there are very few *correct* messages compared to the total amount of possible messages. (As an analogy, think about how there are very few correctly spelled words compared to the total amount of possible ways to arrange a sequence of letters.) The number of incorrect messages is much greater than the number of correct messages. That said, given sufficiently high error rate, and encoded message can become the encoded version of a *different correct message.* Additionally, if the error rate is 100%, then *every single bit gets flipped*. If every single bit gets flipped, then the file maintains a very *specific and predictable relationship* with the original version — it is precisely the inverse. This explains why an 100% error rate does not increase the entropy at all.



Encoding and Entropy

Let's now set the error rate aside, assume the error rate is 0%, and instead focus on how *encoding* influences entropy. How can we expect *encoding* to influence entropy? Considering that encoding a message transforms something readable into a chaotic scramble of completely unreadable characters, it might seem reasonable to predict that encoding would *increase* entropy. But given more thought and observation, we can see that encoding will always *reduce* the amount of entropy. Why is this? When we encode a file, we *double* the number of bits. Meanwhile, we maintain the *pattern* and *uniformity.* Therefore, the amount of entropy *per size of file* is drastically reduced.