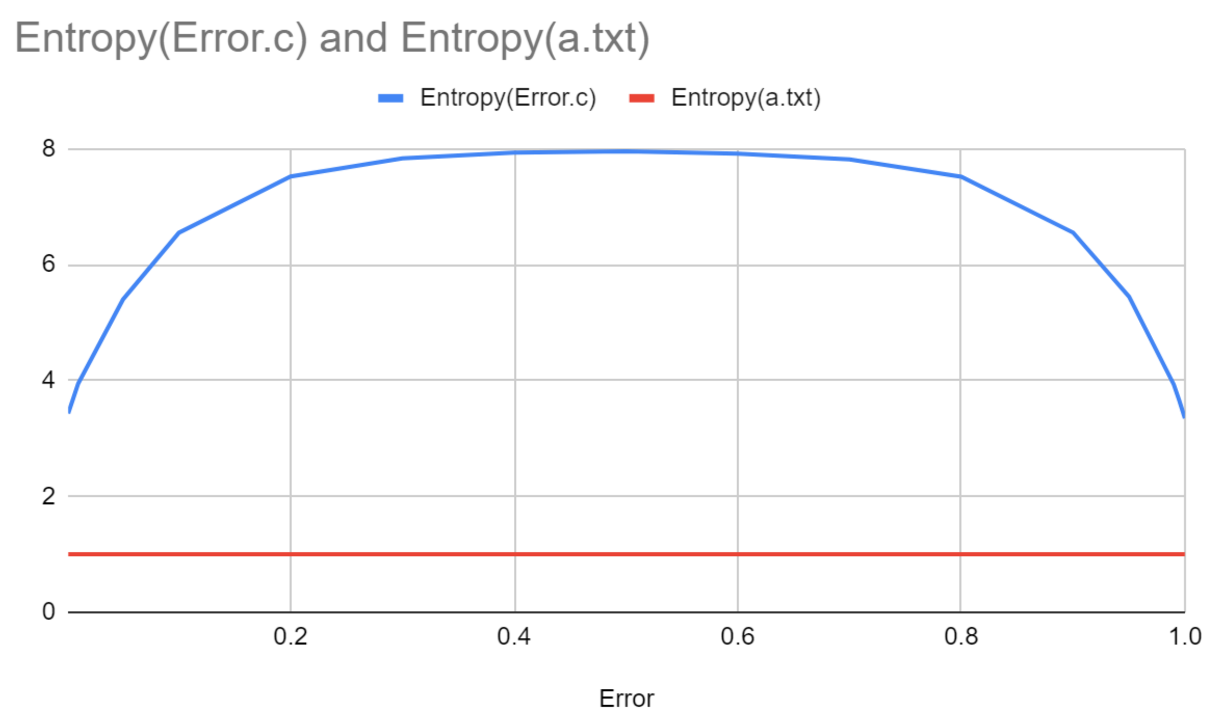
Govind Pillai
gopillai@ucsc.edu
5/9/2021

CSE 13s Spring 2021
Assignment 5: Hamming Codes
Lab Writeup

GRAPH:



Error.c entropy w/o encoding: 4.724378

A.txt entropy w/o encoding: 0.000000

Analysis:

Entropy can be defined as the variation there is in something. Claude Shannon defines it as the amount of information that is produced by some process. In my graph, I ran 2 files against the same error rate and seed values, and then plotted their entropy. For the file “error.c”, the entropy with a 0.001 error rate is around 3.431. Then the graph makes a bell curve-like graph back to the same value when run against a 100% error rate. I was surprised when I ran the file “a.txt” to find that the entropy was 1 for all the data points I ran against “error.c”. Going back to the definition of entropy, it is a measure of uncertainty. Changing the error rate for a file with just one byte, is not going to change its uncertainty, that is why the entropy stays constant for this file. Even a text file with a lot of “a” characters who produce an entropy of 1 because there is no uncertainty. All of the characters are ‘a’. “Error.c” is a different story. With 3872 different bytes being processed, the error rate does have an effect on its entropy (or uncertainty). Based on the graph, the entropy grows with the error rate until the rate hits 0.5. Then it climbs back down to its original value. This is because codes can decode to other messages. When you implement an error rate of more than 50% they decode to the same messages that an entropy of 0-49% would. This is why we get the bell curve shape.