Winston Yi
wyi10@ucsc.edu
4/24/2021

CSE13s Spring 2021
Assignment 5: Hamming Codes
Writeup Document

Encode:

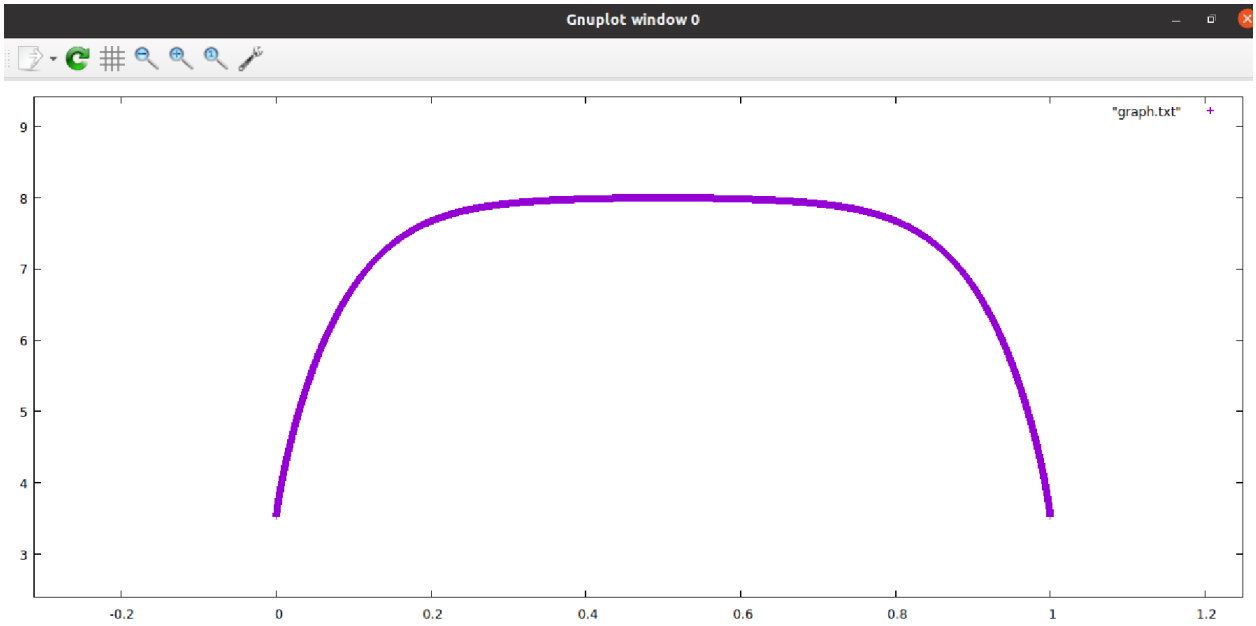

Figure 1: Entropy after encoding with respect to error rate. File measured was random.txt.

The entropy of a file after encoding it supports a nice, albeit thick, bell curve where the entropy seems to flatten at an error rate from (0.4, 0.6). This is likely because at these levels of error rates the program will output the most amount correct errors and not match errors to a ham status in the lookup table. At the error rate of 50% (0.5), it is the most likely that half the bits in each byte have been flipped, which causes the maximum entropy because when 3 bits are flipped in each byte it matches to an incorrect look table which then decodes successfully but to an entirely different message which still lowers the entropy.

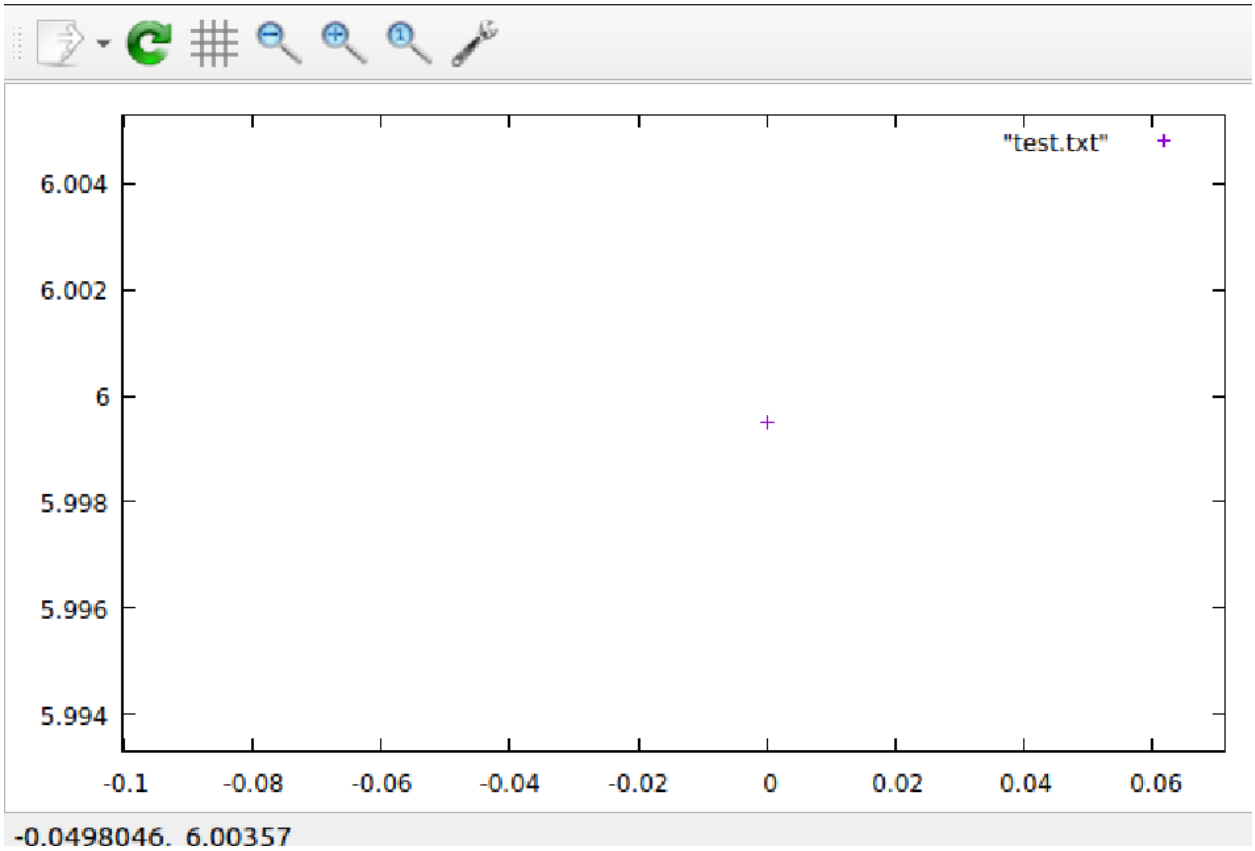

Figure 2: Entropy of the file without any operations done

What is interesting to note here is that the entropy of the original file is less than the encoded version of the file. I believe this to be because the entropy of the file is likely the limit for the ascii characters which can only have 128 possible options. However, when encoded into other random options, the possibilities of what each character, or byte, increases because we double the number of bytes and are no longer bound by the limit of ASCII characters. The difference between the file encoded and just the original seems to be rather large, as the entropy for the encoded file is approximately 3.5 as seen in Figure 1 where the error rate is 0, and the entropy for the original file is approximately 6 as seen in Figure 2.

Conclusion:

Overall, the most interesting thing to note was the level of entropy and how it formed a bell curve. The reason for the horizontal asymptote at x = 0.5 was because an error rate of 0.5% is the perfect point to have the maximum entropy as the encoder will detect an error and be unable to correct it (3 erroneous bits is regarded as no error). A greater or lesser error rate will cause the encoder to match the lookup table regardless of whether or not the message matches the original.