IT – HW1

Tiago Filipe Sousa Gonçalves

March 31, 2021

Exercise 6

Consider a binary Huffman code constructed for symbols $s_1, s_2, ..., s_q$ with probabilities $p_1, p_2, ..., p_q$. Suppose that $p_1 > 0.5$. Must the codeword for s1 be only one bit long? Explain why this must be true or give an example showing that it is not always true.

Let's recall two important properties of the a binary Huffman code: 1) the two least probable letters have codewords with the same maximum length (l_{max}) and 2) given any two letters a_j and a_k, if $P[a_j] >= P[a_k]$, then $l_j <= l_k$, where l_j is the number of bits in the codeword for a_j. Taking this into account, if s_1 is the symbol with the highest probability (since $p_1 > 0.5$), we may assume that it will be always the last symbol to be added to the representation tree (*i.e.*, it will always be on top of the tree). Hence, it is reasonable to assume that this must always be true.

Exercise 8

What should be the mutual information between the plaintext and ciphertext in a perfect cryptographic system?

In a perfect cryptographic system, the mutual information between the plaintext and the ciphertext should be equal to zero.

Exercise 9

Read about the substitution cipher. What are the main properties of the plaintext that leak to the ciphertext?

Let's start by recalling that each language has certain features, such as the frequency of letters (or of groups of two or more letters). Let's also assume that, in this exercise, we are referring to the monoalphabetic substitution cypher. Therefore, we may assume that each letter in the ciphertext corresponds to only one letter in the plaintext letter. Therefore, although the message is built with different letters, the language features will be preserved (i.e., the frequency distributions of symbols in the plaintext and the ciphertext are identical, as well as any structure or pattern in the plaintext is preserved intact in the ciphertext). This makes substitution cyphers more vulnerable to frequency analysis attacks.