

Homework Assignment VIII

Reading Assignment: Kuc Chapter 7

- Let's explore the encoding of a binary source with skewed statistics: $P[1] = \frac{1}{8}$ and $P[0] = \frac{7}{8}$.
 - What is the entropy \mathcal{H} of this bit-stream?
 - What is the average code-word length for the fixed-length code in this case? What is the average code-word length if we apply Huffman code directly?
 - One solution to this problem is to combine two consecutive binary symbols to form a new symbol. Find the probabilities of occurrence for the "new" symbols: $P[00]$, $P[01]$, $P[10]$, $P[11]$.
 - What is the entropy of the modified source in Part **c**? How does it compare to the original entropy?
 - Construct the Huffman code for the modified source of Part **c**. Show the Huffman tree as well as the Huffman code table.
 - What is the average Huffman code-word length now and how does it compare to the two entropy figures computed earlier?
 - Encode the following bit-stream using your Huffman code:

0001000010010000100000

How many bits are you able to save?

- Consider the coding process with variable-length codes of the student-grade data $\{ A, B, C, D, F \}$ whose source statistics are: $P[A] = 0.5$, $P[B] = 0.2$, $P[C] = 0.2$, $P[D] = 0.05$, and $P[F] = 0.05$.
 - What is the entropy of this source? How many bits do we need to represent each symbol using a fixed-length code?
 - Design the Shannon-Fano code. Construct the code table as well as the code tree. What is the average code-word length?
 - Design the Huffman code. Construct the code table as well as the code tree. What is the average code-word length?
 - The source statistics now change to: $P[A] = 0.75$, $P[B] = 0.08$, $P[C] = 0.07$, $P[D] = 0.05$, and $P[F] = 0.05$. Re-design the Shannon-Fano code. Construct the code table as well as the code tree. What is the average code-word length?
 - Re-design the Huffman code for the source above. Construct the code table as well as the code tree. What is the average code-word length?
 - What happens if we use the old Huffman code to encode the new data set in Part (d) and use the new Huffman code in Part (d) to encode the old data set? What are the resulting two average code-word lengths? Comment on your results.

3. Consider the encoding process with variable-length codes of the message

ELECTRICALENGINEERINGISGREAT

- (a) What is the estimated entropy of the message? How many bits do we need to represent each symbol using a fixed-length code?
- (b) Design the Shannon-Fano code. Construct the code table as well as the code tree.
- (c) Find the final encoded bit stream with your Shannon-Fano code. How many bits did you spend to encode the message?
- (d) Design the Huffman code. Construct the Huffman table as well as the Huffman tree.
- (e) Find the final encoded bit stream with your Huffman code. How many bits did you spend?
- (f) If there is a single bit error in the Huffman stream, will the error affect one letter in the decoding process or more? Justify your answer by an example.

4. Consider the encoding process with variable-length codes of the following DNA sequence

AAATCCGTAGCAAACA

- (a) What is the entropy of this DNA sequence?
- (b) Design the Shannon-Fano code. Construct the code table as well as the code tree. What is the average code-word length?
- (c) Design the Huffman code. Construct the code table as well as the code tree. What is the average code-word length?
- (d) The Morse code of the 4 symbols *ACTG* is shown below. What is the average code-word length? Can you come up with the Morse code tree? What problem can you observe from the Morse tree?

A ● — **C** — ● — ●
T — **G** — — ●

- (e) Compare the efficiency of the three different VLC codes: Shannon-Fano, Huffman, and Morse in this case. What is the best code and how does it compare to the entropy?

Due date: **November 28** in class