

Written Exam in
Data compression
TSBK08

8th June 2023 8:00 - 12:00

Location:	U3
Examiner:	Harald Nautsch
Teacher:	Harald Nautsch, 1361
Department:	ISY
Module:	TEN1
Number of problems:	7
Number of pages:	4
Permitted equipment:	Calculator, general English dictionaries
Other:	Answers can be given in English or in Swedish. The teacher will visit at around 9:15 and 10:45
Grades:	0-13 U 14-19 3 20-25 4 26-30 5

- 1
- a) What are the properties of a source that makes it possible to compress its output?
(2 p)
- b) Describe the following classes of source codes and how they relate to each other.
- Uniquely decodable codes.
 - Instantaneous codes.
- (2 p)
- c) Explain what a Golomb code is and what type of probability distribution it is good for.
(2 p)
- d) Explain how prediction with partial match (ppm) coding works.
(2 p)

- 2
- Assume that we have a stationary source X_i and that we code n symbols at a time using an optimal code. Show that the mean data rate R of the code approaches the entropy rate of the source when n grows to infinity.
(4 p)

- 3
- A second order Markov source X_i with alphabet $\mathcal{A} = \{a, b\}$ is given by the transition probabilities $p(x_i|x_{i-1}x_{i-2})$ (note the symbol order) below

$$\begin{aligned}
 p(a|aa) &= 0.9, & p(b|aa) &= 0.1 \\
 p(a|ab) &= 0.8, & p(b|ab) &= 0.2 \\
 p(a|ba) &= 0.4, & p(b|ba) &= 0.6 \\
 p(a|bb) &= 0.1, & p(b|bb) &= 0.9
 \end{aligned}$$

- a) Calculate the entropies $H(X_i)$, $H(X_i, X_{i+1})$ and $H(X_i|X_{i-1}, X_{i-2})$ for the source.
(3 p)
- b) Construct a Huffman code for triples of symbols and calculate the rate of the code (in bits/symbol).
(3 p)

- 4 A memoryless source has the alphabet $\mathcal{A} = \{a, b, c\}$ with symbol probabilities $p(a) = 0.7$, $p(b) = 0.2$ and $p(c) = 0.1$. Show how arithmetic coding works by coding the sequence

caaaba

Give both the corresponding interval and the codeword. You can assume that all calculations are exact.

(4 p)

- 5 A source has the alphabet $\mathcal{A} = \{n, o, p, q, r, s, t, u\}$.

We want to code the source using LZSS.

- a) Assume that we want to use a search buffer size of 32 and that we want use 3 bit fixed length codewords to code the match lengths. What is the shortest match length that should be coded as a match instead of a sequence of single symbols?

(1 p)

- b) Code the sequence beginning with

turturtususuportusususppppportn...

Give the resulting binary codewords.

(3 p)

- 6 A source has the alphabet $\mathcal{A} = \{s, t, u, v\}$.

Code the sequence

susstuss

using BWT (Burrows-Wheeler's transform) followed by mtf (move-to-front coding). The block size should be 8.

(3 p)

- 7 Calculate the differential entropy of the continuous random variable X with probability density function $f(x)$

$$f(x) = \begin{cases} 0.4 & ; |x| \leq 1 \\ 0.1 & ; 1 < |x| \leq 2 \\ 0 & ; \text{otherwise} \end{cases}$$

(1 p)