LINKÖPING UNIVERSITY

Written Exam in
Data compression
TSBK08

9th June 2022 8:00 - 12:00

Location:	TER1
Examiner:	Harald Nautsch
Teacher:	Harald Nautsch, 0701718715
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 9:15 and 10:45
Grades:	0-13 U 14-19 3 20-25 4 26-30 5

1 a) Describe the following classes of source codes and how they relate to each other.

- Uniquely decodable codes.
- Instantaneous codes.
- Tree codes.

(3 p)

b) Explain how adaptive arithmetic coding works.

(2 p)

c) Explain what a Golomb code is and what type of probability distribution it is good for.

(2 p)

d) Explain what the rate-distortion function is and how it is calculated for a stationary memoryless random source.

(2 p)

2 Formulate Kraft's inequality and give a proof of it.

(4 p)

3 A memoryless source has the alphabet

$$\mathcal{A} = \{a,\ b,\ c,\ d,\ e,\ f,\ g,\ h,\ i,\ j\}$$

The symbol probabilities are

$$p(a) = 0.25,\ p(b) = 0.2,\ p(c) = 0.14,\ p(d) = 0.12,\ p(e) = 0.07$$

$$p(f) = 0.07,\ p(g) = 0.06,\ p(h) = 0.03,\ p(i) = 0.03,\ p(j) = 0.03$$

Construct a Huffman code for the source and calculate the resulting average data rate (in bits/symbol) of the code.

(3 p)

4 A source has the alphabet $\mathcal{A} = \{a,\ b,\ c,\ d\}$. A symbol sequence of length 8 is coded using BWT and mtf. The resulting index is 2 and the mtf-coded sequence is 1,0,0,3,0,2,0,0. Decode the symbol sequence.

(3 p)

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

$$\begin{aligned} p(a|aa) &= 0.9, \quad p(b|aa) = 0.1 \\ p(a|ab) &= 0.6, \quad p(b|ab) = 0.4 \\ p(a|ba) &= 0.3, \quad p(b|ba) = 0.7 \\ p(a|bb) &= 0.2, \quad p(b|bb) = 0.8 \end{aligned}$$

a) Calculate the entropies $H(X_i)$, $H(X_i|X_{i-1})$ and $H(X_i, X_{i+1}, X_{i+2}, X_{i+3})$ for the source.

(3 p)

b) Code the sequence

$$aaabba$$

using arithmetic coding. The coding should use the memory of the source. Give both the interval and the corresponding codeword. You can assume that the source is in state aa when the coding starts and that all calculations are performed with infinite precision.

(4 p)

6 A source has the alphabet $\{r,\ s,\ t,\ u\}$. A sequence from the source is coded using LZW and gives the following index sequence:

$$1, 3, 1, 6, 5, 4, 0, 10, 0, 2, 8, 11, 13, \ldots$$

The starting dictionary is:

index	sequence
0	r
1	s
2	t
3	u

Decode the index sequence. Also give the resulting dictionary.

(3 p)

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

$$f(x) = \begin{cases} 1/2 & ; 0 \leq x \leq 1 \\ 1/3 & ; 1 < x \leq 2 \\ 1/6 & ; 2 < x \leq 3 \\ 0 & ; \text{otherwise} \end{cases}$$

(1 p)