Exercise Sheet 9

Exercise 32   Coding Theory: Shannon-Fano and Huffman Code

Determine

   a) a Shannon-Fano coding scheme/question tree and
   b) a Huffman coding scheme/question tree

for the symbols s_1 to s_6, which occur with the probabilities

$$P(s_1) = 0.05, P(s_2) = 0.10, P(s_3) = 0.12, P(s_4) = 0.20, P(s_5) = 0.25, P(s_6) = 0.28,$$

respectively! Compute the average code length and the code efficiency!
(code efficiency: ratio of Shannon entropy to average code length)

Exercise 33   Decision Trees: Attribute Selection Measures

Compute the information gain and the χ^2 measure for the following two contingency
tables, which refer to two descriptive attributes A, B and one class attribute C!

		A		
		a_1	a_2	a_3
	c_1	9	4	3
C	c_2	3	9	4
	c_3	4	3	9

		B		
		b_1	b_2	b_3
	c_1	9	4	3
C	c_2	6	6	4
	c_3	1	6	9

How may one describe the selection behavior of the two measures intuitively?
(Hint: Mind the first row and the last column of the two tables.)

Exercise 34   Decision Trees: Pruning

Prune the following decision tree using the approach of pessimistic pruning!
(parameter: 0.5 additional errors)

Decision tree:

- att1
  - 1 → att2
    - 1 → c_1 c_2 / 12 1
    - 2 → att3
      - 1 → att4
        - 1 → c_1 c_2 / 2 0
        - 2 → c_1 c_2 / 1 1
        - 3 → c_1 c_2 / 1 2
      - 2 → c_1 c_2 / 5 0
  - 2 → att3
    - 1 → att4
      - 1 → c_1 c_2 / 3 2
      - 2 → c_1 c_2 / 0 7
      - 3 → c_1 c_2 / 0 7
    - 2 → c_1 c_2 / 0 22