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 $\chi^2$ measure for the following two contingency tables, which refer to two descriptive attributes $A, B$ and one class attribute $C$!

<table>
<thead>
<tr>
<th></th>
<th>$A$</th>
<th>$B$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$a_1$</td>
<td>$b_1$</td>
</tr>
<tr>
<td>$c_1$</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>$c_2$</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>$c_3$</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

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)