582487 Data Compression Techniques (Spring 2012)
Exercises 1 (25 January)

Solve the following problems before the exercise session and be prepared to present your solutions at the session.

1. Are the following codes uniquely decodable? Justify your answers.

   (a) a $\mapsto$ 0, b $\mapsto$ 11, c $\mapsto$ 010, d $\mapsto$ 101

   (b) a $\mapsto$ 0, b $\mapsto$ 11, c $\mapsto$ 001, d $\mapsto$ 101

2. In the Huffman code of Example 1.7, some codewords have length 4. Construct a Huffman code (optimal binary prefix code) for the same alphabet and the same probabilities such that no codeword is longer than 3. Verify that the two codes have the same average code lengths. *Hint:* Make different choices during the execution of the Huffman algorithm.

3. Shannon-Fano coding is another method for constructing prefix codes. Find out how it works (e.g. from Wikipedia).

   (a) Apply Shannon-Fano coding to the following probability distribution.

symbol	a	b	c	d	e
prob.	0.38	0.17	0.16	0.15	0.14

   (b) Show that the resulting code is not optimal.

4. The proof of Kraft's inequality in the lecture notes is constructive: it describes an algorithm for constructing a prefix code for any valid set of codeword lengths. Use that algorithm to design a prefix code for the alphabet $\{a, b, c, d, e, f\}$ with the following codeword lengths

symbol	a	b	c	d	e	f
length	4	3	1	3	4	3

   Specify also the intervals associated with the codewords.

5. A prefix code is *canonical* if the codewords in lexicographical order are also in the order of their lengths. For example, the code on the left is not canonical but the one on the right is:

00	00
0100	01
0101	10
011	110
10	1110
11	1111

   Show that for any prefix code there exists a canonical prefix code with the same codeword lengths.

6. Give the codes $\gamma(n)$, $\delta(n)$ and $\mathrm{GR}_4(n)$ for $n \in \{100, 200, 400\}$.