Advanced Topics of Algorithms

Tasks 2, 17.2.2016

1. Consider a statistical model of a binary image, in which $p_W = 0.75$ and $p_B = 0.25$. We have a pixel sequence (W, W, W, B, B, W, W, W) to be compressed by arithmetic coding. Which interval this sequence corresponds to, and what is the actual code (bit sequence) of this code interval?

2. Assume that *Academic Better Company* ABC offers the following odds. For example, if you predict that Arsenal wins and it happens, you will get your money back with 3.8 times (1 € → 3.8 €).

-	Arsenal – Bayern München:	1=3.80	X= 3.80	2=1.85
-	Student will pass DAA exam:	Yes=1.50	No=3.00	
-	Sun will rise tomorrow morning:	Yes=1.01	No=101	
-	You have answered all correct here:	Yes=3.00	No=4.00	Half=5.00

(a) Conclude the probabilities for each of the events happening.
(b) Calculate the entropies for each output. Explain the difference of compression and betting.
(c) Which one of the above betting would you put your money in?
(d) What if you **know** the real probabilities (e.g. Arsenal wins 25% and Bayern 50%). Would you bet?
(e) Discuss about possible strategies for professional gambler.

3. Take at least two Finnish, two English and two another language web page (or just text file, you can omit spaces and delimiters). Calculate the frequencies and entropies of of all letters 'a' to 'ö' and sort them in the decreasing order according to the frequencies. Does the different language have different entropies?

4. Same as in previous task, but use 1st order context model where the previous char is the context.

5. Cross-entropy calculates the code length if the data A is coded using a probability model obtained from another data B. It describes the inefficiency of the model B for coding data A. It is always greater or bigger than the entropy using the correct model obtained directly from data A. Calculate the cross-entropies for the above language detection tasks.

6. Study K-means clustering algorithm using these two material:

```
http://cs.uef.fi/sipu/clustering/animator/
http://cs.uef.fi/pages/franti/cluster/Clustering-Part2a.ppt
```