AI2 Module 3
Tutorial 1: Sample Solutions

Jacques Fleuriot School of Informatics
(Amended by David Talbot January 6, 2005.)

The main aim for this tutorial is to gain familiarity with the entropy equations and the idea of "information" counted in bits. The last part is important as it relates to the splitting stage of the decision tree learning algorithm.

1. We need 2 bits per marble: $1024 \cdot 2 = 2048$

2. The following encoding is *one* possibility:

   Use code: (Red,0), (Green,10), (Blue,110), (Yellow,111).
   We need $512 \cdot 1 + 256 \cdot 2 + 2 \cdot 128 \cdot 3 = 1792$ bits which is an improvement.

   Yes we can drop the commas as our code is "prefix free" - that is no codeword is a prefix of another codeword (would happen e.g. if we used 0, and 01 as codewords). So, we can scan from left to right and stop as soon as we identify a codeword.

   (Note that we ignore the code table. This is reasonable as the table is small but we code many marbles.)

3. Everything is a power of 2 so computing log is easy:

$$\begin{aligned} I &= \frac{512}{1024}\log_2\frac{1024}{512} + \frac{256}{1024}\log_2\frac{1024}{256} + \frac{128}{1024}\log_2\frac{1024}{128} + \frac{128}{1024}\log_2\frac{1024}{128} \\ &= \frac{1}{2}\log_2 2 + \frac{1}{4}\log_2 4 + 2\frac{1}{8}\log_2 8 = \frac{1}{2} + \frac{1}{2} + 2\frac{3}{8} = 1.75 \end{aligned}$$

   The total number of bits required is $1024 \cdot 1.75 = 1792$. So our encoding achieves the entropy bound exactly. This is not always possible (e.g. if one doesn't have nice numbers for the probabilities).

4. The main thing is that we can use less space *because* of the extra information we have (the regularity in data). This is a real saving in our case but will not always work.

   Using I_1 for part 1 and I_2 for part 2:

$$\begin{aligned} I_1 &= \frac{512}{600}\log_2\frac{600}{512} + \frac{88}{600}\log_2\frac{600}{88} = 0.601 \\ I_2 &= \frac{256}{424}\log_2\frac{424}{256} + \frac{128}{424}\log_2\frac{424}{128} + \frac{40}{424}\log_2\frac{424}{40} = 1.282 \\ \text{Total} &= 600 \cdot 0.601 + 424 \cdot 1.282 = 904.2 \end{aligned}$$

   NB: This is not a code but just what the entropy tells us (about a lower bound for average code length).

   Coding: For part 1 we just have 2 colours; use 0 and 1 to get a total of 600 bits. For part 2 use: (G,0), (B,10), (Y,11) for a total of $256 \cdot 1 + 128 \cdot 2 + 40 \cdot 2 = 592$. The grand total of the two parts is 1192 bits; an improvement on the original.

   Note that the main "inefficiency" left is in part 1 where we use 600 bits compared with $600 \cdot 0.601$.