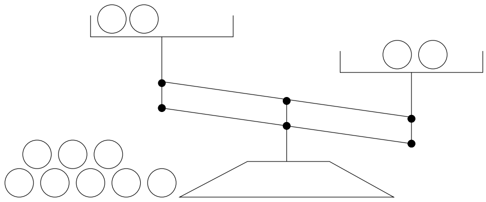
Information Theory, Pattern Recognition and Neural Networks

Part III Physics, January 2008

Please work on the following exercise before lecture 2.

The weighing problem



You are given 12 balls, all equal in weight except for one that is either heavier or lighter. You are also given a two-pan balance to use. In each use of the balance you may put any number of the 12 balls on the left pan, and the same number on the right pan, and push a button to initiate the weighing; there are three possible outcomes: either the weights are equal, or the balls on the left are heavier, or the balls on the left are lighter. Your task is to design a strategy to determine which is the odd ball *and* whether it is heavier or lighter than the others *in as few uses of the balance as possible.*

You may be used to receiving photocopied handouts containing lecture notes, exercises, and worked solutions. For this course all these handouts have been combined into a single document, called a *book*. The course textbook is *Information theory, inference, and learning algorithms*, by David J.C. MacKay. Cambridge University Press (2003). (Rayleigh library: 39 M 20.) This 640-page textbook covers the whole course, and a whole lot more. I encourage all students to buy or borrow this textbook. The book's cover price is £35; if you buy it at the CUP bookshop with University ID, it costs £28. You can also buy it from `amazon.co.uk` for £33.25. But **if you don't want to buy it, there are alternative options:**

1. You can **borrow the book** from me: I have copies available for a returnable cash deposit of £21.

2. The book is also available for **free download** from

   `http://www.inference.phy.cam.ac.uk/itprnn/`.

   You can download the whole thing, or individual chapters.

3. **Money-back Guarantee:** *If you buy the book, then decide that you don't want to keep it, I will buy it from you for a good price and sell it on to a future student.*

Note-taking: During lectures, I encourage you to take notes, rather than just follow along in the textbook. Learning is an active process, not a passive one. I aim to lecture at the right pace so that you can take notes and understand. If the pace is wrong, please let me know.

Exercises: The book contains numerous exercises, complete with worked solutions. A subset of these exercises will be designated as the exercises you should work on before supervisions. [Generally, these exercises are the ones marked 'highly recommended' by a marginal rat.] *I encourage you not to look at the worked solutions for these exercises before you have attempted them yourself.*

The first bunch of exercises are: 1.3 (p. 8), 1.5–7 (p. 13), 1.9, & 1.11 (p. 14). The first supervision will focus on exercise **1.9** (p. 14).

Supervisions: Officially, you are meant to get three supervisions in small groups. However, feedback from students indicates a preference for *six* supervisions in larger groups. The supervisions will be weekly on Mondays at 2pm or 4pm starting **Monday 4th February**.

Location of 2pm supervisions: TCM Seminar Room, Mott building.

Location of 4pm supervisions: Mott seminar room, Mott building (Room 531).

Please attend one supervision per week, and sign the attendance register each time.

Your feedback on all aspects of the course is welcomed.

Feedback given *early* is more valuable than feedback delivered through the end-of-term questionnaires. Early feedback allows any problems arising to be fixed immediately rather than next year!

Course Synopsis

Introduction to information theory. The possibility of reliable communication over unreliable channels. The (7,4) Hamming code and repetition codes.

Entropy and data compression. Entropy, conditional entropy, mutual information, Shannon information content. The idea of typicality and the use of typical sets for source coding. Shannon's source coding theorem. Codes for data compression. Uniquely decodeable codes and the Kraft–MacMillan inequality. Completeness of a symbol code. Prefix codes. Huffman codes. Arithmetic coding.

Communication over noisy channels. Definition of channel capacity. Capacity of binary symmetric channel; of binary erasure channel; of Z channel. Joint typicality, random codes, and Shannon's noisy channel coding theorem. Real channels and practical error-correcting codes. Hash codes.

Statistical inference, data modelling and pattern recognition. The likelihood function and Bayes' theorem. Clustering as an example.

Approximation of probability distributions. Laplace's method. (Approximation of probability distributions by Gaussian distributions.) Monte Carlo methods: Importance sampling, rejection sampling, Gibbs sampling, Metropolis method. (Slice sampling, Hamiltonian Monte Carlo, Overrelaxation, and exact sampling.*) Variational methods and mean field theory. Ising models.

Neural networks and content-addressable memories. The Hopfield network.

The roadmap on page vii of the book indicates which chapters are relevant to this course.

Errata for the fourth printing of the textbook

p.214 Table should read $p_b = \frac{3}{N}\binom{N}{2}f^2$: probability of bit error to leading order (not block error)

p.223 Soln to ex 13.4. should read: The probability of block error to leading order is $p_{\rm B} = \binom{N}{2}f^2$. The probability of bit error to leading order is $p_{\rm b} = \frac{3}{N}\binom{N}{2}f^2$.

p.461 Near bottom of page and in caption of figure 37.4: Replace $p_{A+} < 10p_{B+}$ by $p_{A+} < 0.1p_{B+}$

p.533 Equation 44.12 right hand side has a sign error. $P(t)/P(t\,|\,y)$ should be $P(t\,|\,y)/P(t)$.


```
David J.C. MacKay   email: mackay@mrao.cam.ac.uk
                      www: http://www.inference.phy.cam.ac.uk/mackay/
tel: 339852          room: 518 Mott Building
```