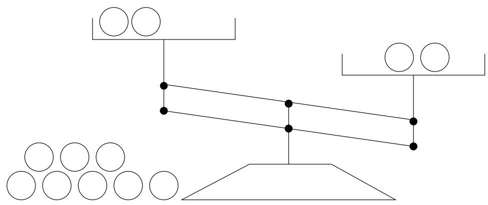
Information Theory, Pattern Recognition and Neural Networks

Part III Physics, January 2005

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.*

Textbook

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. All students are encouraged to buy this textbook. The book is also available for free on-screen viewing at

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

I surveyed the 2003 class to find out what they would be willing to pay for the book. The median answer was £25. The book's cover price is £30; if you buy it at the CUP bookshop with University ID, it costs £24. **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. Assuming that you would like this, the supervisions will be on Thursdays at 3pm or 4.30pm starting in week 3. (**Thursday 3rd February**.) Location: Ryle seminar room, Rutherford building (Room 930). Please attend one supervision per week, and sign the attendance register each time.

P.T.O.

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 (details of proof involving H_δ are not examinable). 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. Inference of discrete and continuous parameters. 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.

Data modelling with neural networks. Interpolation and classification using a single neuron. (Multilayer perceptrons*). Backpropagation algorithm. Learning algms. viewed in terms of inference.

Neural networks as information storage devices. Capacity of a single neuron. Hopfield network and its relationship to spin glasses. Hopfield network for optimization problems, e.g., travelling salesman problem. Boltzmann machine. Hopfield network as a mean field approximation to the Boltzmann machine. Boltzmann machine learning algorithm.*

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

Errata for the third printing of the textbook

p.3: The word "communication" is missing from the first line of Chapter 1.
The fundamental problem of communication is that of ...

p.30: before Ex 2.8 replace $<$ by $\leq$ so it reads: $P(v) \leq 1$ for any interval (a, b).

p.36: Ex 2.17: "what is b as a function of p?" should be "what is p as a function of b?"

p.42: eq 2.68: Replace a by b.

p.80: eq 4.28. Second "$\simeq$" should be "$=$"

p.301: Insert minus sign thus after eq 22.7: ... "which is $-1/\sigma_\mu^2$," ...

p.598-9: Gamma function section: all "log" should be "ln".

David J.C. MacKay email: mackay@mrao.cam.ac.uk
www: http://www.inference.phy.cam.ac.uk/mackay/
tel: 339852 room: 961 Rutherford Building