Information Theory, Pattern Recognition and Neural Networks

HANDOUT 3 MARCH 11, 2005

1 For the final supervision

On **Wednesday 16th March**, hand in suggestions for topics to cover. (Four exercises are listed below.)

2 Help!

Next year, the department are forcing all minor options to shrink down by 25%, from 16 lectures to 12.

I welcome your advice. (a) Which topics do you think should definitely be kept? (b) Which topics would you recommend cutting?

3 Course summary: central chapters

Data compression and noisy channel coding (Chapters 1–6, 8–10, 14).
Inference and data modelling. (Chapters 20 and 22).
Monte Carlo methods (Chapter 29).
Variational methods (Chapter 33).
Laplace's method (Chapter 27).
Ising models (Chapter 31).
Neural networks (Chapters 38, 39, 42, and some of 41 and 44).

4 Related chapters

Given what you've learned already, I expect you'll find these chapters accessible and interesting.

Ch 12 (Hash codes).

Ch 28 (Model comparison and Occam's razor).

Chs 30 and 32 (Efficient and Exact Monte Carlo methods).

Ch 40 (Capacity of a single neuron).

Chs 47 and 50 (Sparse graph codes).

5 Exercises that have been recommended

1: 1.3 (p.8), 1.5-7 (p.13), 1.9, & 1.11 (p.14).
2: ex 5.29.
5.22, 5.27, 6.7, 6.15, 6.17.
then if you need more practice, 5.26, 5.28, 15.3 (p.233)
3: 9.17 (p.155) 10.12 (172) 15.12 (235)
then if you need more practice, 15.11, 15.13, 15.15.
4: Ex 22.5.
Ex 3.10 (p57). 8.10. 9.19. 9.20. 15.5, 15.6, (233) 22.8. Additional suggestions: 8.3 (140), 8.7, 22.4.
5: HANDOUT 2. Exercise 27.1. 29.14, 33.5, 33.7.
6: 22.11 (Make sketches of the posterior distribution of the sailor).
39.4, 39.5 (single neuron).

Two approaches have been suggested to the problem in handout 2:

Bayes' theorem, in which the log likelihood ratio is

$$\log \frac{P(\mathbf{x} \mid \mathcal{H}_P)}{P(\mathbf{x} \mid \mathcal{H}_Q)} = \sum_i F_i \log \frac{p_i}{q_i},$$

where i runs over characters in the alphabet, F_i is the number of times character i actually occurred in the data string $\mathbf{x}$, and the two models $\mathcal{H}_P$ and $\mathcal{H}_Q$ state that the symbols come i.i.d. from the distributions $\mathbf{p}$ and $\mathbf{q}$ respectively.

Chi-squared. In a chi-squared approach, we compute the two measures of goodness of fit,

$$\chi_P^2 = \sum_i \frac{(F_i - p_i N)^2}{p_i N}$$

$$\chi_Q^2 = \sum_i \frac{(F_i - q_i N)^2}{q_i N},$$

where N is the number of characters received; then go for the hypothesis with smaller χ^2.

These two approaches do not always make the same decision. (Notice that the log likelihood ratio is a *linear* function of $\{F_i\}$ whereas $\chi_P^2 - \chi_Q^2$ has a *quadratic* dependence on $\{F_i\}$.)

Task: seek out examples that magnify the differences between these two approaches. (a) Can you find an example data set, and pair of hypotheses $\mathbf{p}$ and $\mathbf{q}$, for which the two approaches give completely different answers? (b) Can you find two data sets that are intuitively equivalent from the point of view of comparing $\mathbf{p}$ and $\mathbf{q}$, but for which one of the approaches gives different answers?

6 What's examinable

TOPICS THAT ARE ALWAYS EXAMINED

Data compression. Evaluating entropy, conditional entropy, mutual information. Symbol codes. Huffman algorithm.

Noisy channels. Evaluating capacity.

Inference problems. Inferring parameters. Comparing two hypotheses. Sketching posterior distributions. Finding error bars.

SOMETIMES EXAMINED

Data compression Typicality. Arithmetic coding.

Noisy channels. Channel coding theorem.

Monte Carlo methods. Metropolis method, Gibbs sampling.

Clustering.

Neural networks. Feedforward networks. Feedback networks.