Information Theory, Pattern Recognition and Neural Networks

HANDOUT 2 FEBRUARY 3, 2009

1 Course summary: central chapters that you should read

Data compression and noisy channel coding (Chapters 1–6, 8–10, 14). (But omitting section 6.4 and 10.4–10.8)
Inference and data modelling. (Chapters 3, (20), 21, and 22; also the Taylor expansion of chapter 27 (p. 341)). (20 isn't covered in class, but may be helpful reading.)

2 Exercises for supervisions

Supervision 1: Invent a code. 1.3 (p.8), 1.5–7 (p.13), **1.9**, & 1.11 (p.14).

Supervision 2: Invent a compressor. Invent a compressor and uncompressor for a source file of $N = 10\,000$ bits, each having a probability $f = 0.01$ of being a `1`. Make use of Huffman codes if you wish, or use other methods. Implement your compressor and uncompressor and/or estimate how well your method works. The website has implementations of the Huffman algorithm in various languages, and a bent coin file as a compression benchmark.

Other exercises: 5.22 (p. 102), 5.27. Then if you need more practice, 5.26, 5.28.

Supervision 3: 5.31, 6.3, 6.7, 6.17. 2.25 (p. 37), 2.26, 2.28. Then if you need more practice, 6.15, 6.18, 15.3 (p. 233).

Supervision 4: 9.17 (p. 155) 10.12 (p. 172) 15.12 (p. 235); then if you need more practice, 15.11, 15.13, 15.15.

Invent a channel. Invent a channel to pose to your colleagues, in the form "what's the capacity of *this*?"

Supervision 5: See 'spy' question below.

Examples 22.1–4 (p. 300) and exercise 22.8.

Ex 3.10 (p. 57); 8.10; 9.19; 9.20; 15.5, 15.6 (p. 233); 8.3 (p. 140), 8.7; 22.11; 22.5.

The spy question. A spy would like you to write a computer program that recognises, given a small number of consecutive characters from the middle of a computer file, whether the file is an English-language document. Assuming that the two alternative hypotheses are that the file is an English-language document ($\mathcal{H}_E$), or that it is a random string of characters drawn from the same alphabet ($\mathcal{H}_R$), describe how you would solve this problem.

Estimate how many characters your method would need in order to work reasonably well.

Maybe you would enjoy writing a program that implements your method?

How would your answers differ if instead the task were to distinguish

(a) English from German?

(b) English from Hsilgne (backwards English)?

[When I say German, let's assume German with no accent characters.]

(Some facts about English and German are supplied in table 1.)

	A	B	C	D	E	F	G	H
English (**e**)	.07	.01	.03	.03	.10	.02	.02	.05
German (**g**)	.06	.02	.03	.04	.15	.01	.03	.04
	I	J	K	L	M	N		
English (**e**)	.06	.001	.006	.03	.02	.06		
German (**g**)	.07	.002	.01	.03	.02	.08		
	O	P	Q	R	S	T	U	
English (**e**)	.06	.02	.0009	.05	.05	.08	.02	
German (**g**)	.02	.007	.0002	.06	.06	.05	.04	
	V	W	X	Y	Z	-		
English (**e**)	.008	.02	.002	.01	.0008	.17		
German (**g**)	.006	.02	.0003	.0003	.01	.14		

Table 1: Letter frequencies of English and German. The entropies of these two distributions are $H(\mathbf{e}) = 4.1$ bits; $H(\mathbf{g}) = 4.1$ bits; and the relative entropies between them are $D_{\mathrm{KL}}(\mathbf{e}||\mathbf{g}) = 0.16$ bits and $D_{\mathrm{KL}}(\mathbf{g}||\mathbf{e}) = 0.12$ bits. The relative entropies between the uniform distribution $\mathbf{u}$ and the English distribution $\mathbf{e}$ are $D_{\mathrm{KL}}(\mathbf{e}||\mathbf{u}) \simeq 0.6$ bits and $D_{\mathrm{KL}}(\mathbf{u}||\mathbf{e}) \simeq 1$ bits.

Supervision 6: See 'how well calibrated' question overleaf.

Invent a supervision. Design an activity for use in a supervision; email your solution to `djcm1@cam`.

3 How well calibrated are your estimates of uncertainty?

Give a 94% confidence interval for the following quantities. Give the tightest interval you can, while remaining 94% sure that the true value is in the interval. *Don't look up answers before you have written down your interval* – the aim of this exercise is to get a feel for how well calibrated your intervals are.

	Quantity	Lower bound	Guess	Upper bound
1	Mass of the textbook (g)			
2	Population of Britain (census, 2001)			
3	Population of Turkey (July 2004)			
4	Population of Luxembourg (July 2004)			
5	Number of British MEPs			
6	Starting pay of University Lecturer (Aug 2004)			
7	Parliamentary salary of MP (1/4/2005)			
8	Council tax, South Cambs. (£/house/yr) (band D, 2005-6)			
9	Fraction of central government expenditure that goes to 'Defence' (2004)			
10	UK prison population (as fraction of whole) (March 2005)			
11	Number of USA nuclear warheads (Feb 2003)			
12	Distance to sun (miles) (on 10 March 2005)			
13	Mean radius of earth (km)			
14	Speed of light ($\mathrm{m\,s^{-1}}$)			
15	Density of Gold ($\mathrm{g\,cm^{-3}}$)			
16	The ratio $\frac{\text{Density of Uranium}^{238}}{\text{Density of Gold}}$			

4 What's on the exam

Data compression. Evaluating entropy, conditional entropy, mutual information. Symbol codes. Huffman algorithm. 'How well would arithmetic coding do?' Applying information theory concepts.

Noisy channels. Evaluating conditional entropy, mutual information. Definition of capacity. Evaluating capacity. Finding optimal input distributions. Inference of input given output. Connection to reliable communication.

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

Past exam questions

The following exercises from the book were exam questions. The **bold** questions are especially recommended (and were all recommended exercises already). Further past exam questions are on the website.

Source coding	Noisy channels	Inference
5.27 ++	10.12	**22.5**
5.28	15.11	**22.8** ++
5.29	15.12	27.1
6.9	15.13	
6.15	**15.15** ++	
6.17		
6.18		
15.3 ++		