Information Theory, Pattern Recognition and Neural Networks (April 2001)

Part III Physics exam 2001

1: Write an essay on one of the following topics.

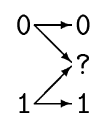
(a) Content-addressable memory. [20]

(b) A critical review of the Metropolis method and Gibbs sampling, emphasising the strengths and weaknesses of these Monte Carlo methods. [20]

2: Answer all four parts.

(a) A binary erasure channel with input x and output y has transition probability matrix:

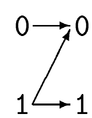
$$Q = \begin{bmatrix} 1-q & 0 \\ q & q \\ 0 & 1-q \end{bmatrix}$$



Find the *mutual information* $I(X;Y)$ between the input and output for general input distribution $\{p_0, p_1\}$, and show that the *capacity* of this channel is $C = 1 - q$ bits. [5]

(b) A 'Z channel' has transition probability matrix:

$$Q = \begin{bmatrix} 1 & q \\ 0 & 1-q \end{bmatrix}$$



Show that, using a (2,1) code, **two** uses of a Z channel can be made to emulate **one** use of an erasure channel, and state the erasure probability of that erasure channel. Hence show that the capacity of the Z channel, C_Z, satisfies $C_Z \geq \frac{1}{2}(1-q)$ bits. [4]

(c) A (7,4) Hamming code is used to communicate over a binary symmetric channel with noise level $f = 0.01$. Estimate (to one decimal place) the *block error probability* of the code. [4]

(d) A (3,1) code consists of the two codewords $\mathbf{x}^{(1)} = (1,0,0)$ and $\mathbf{x}^{(2)} = (0,0,1)$. A source bit $s \in \{1,2\}$ having probability distribution $\{p_1, p_2\}$ is used to select one of the two codewords for transmission over a binary symmetric channel with noise level f. The received vector is $\mathbf{r}$. Show that the posterior probability of s given $\mathbf{r}$ can be written in the form

$$P(s=1|\mathbf{r}) = \frac{1}{1 + \exp\left(-w_0 - \sum_{n=1}^{3} w_n r_n\right)},$$

and give expressions for the coefficients $\{w_n\}_{n=1}^{3}$ and the bias, w_0. [5]

Describe, with a diagram, how this optimal decoder can be expressed in terms of a 'neuron'. [2]

3: Answer both parts.

(a) A binary source X emits independent identically distributed symbols with probability distribution $\{f_0, f_1\}$, where $f_1 = 0.01$. Find an optimal uniquely-decodeable symbol code for a string $\mathbf{x} = x_1x_2x_3$ of **three** successive samples from this source. [4]

Estimate (to one decimal place) the factor by which the expected length of this optimal code is greater than the entropy of the three-bit string $\mathbf{x}$. [2]

[$H_2(0.01) \simeq 0.08$, where $H_2(x) = x \log_2(1/x) + (1-x)\log_2(1/(1-x))$.]

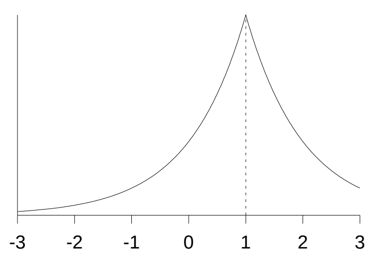
An *arithmetic code* is used to compress a string of 1000 samples from the source X. Estimate the mean and standard deviation of the length of the compressed file. [6]

(b) In an experiment, the measured quantities $\{x_n\}$ come independently from a bi-exponential distribution with mean μ,

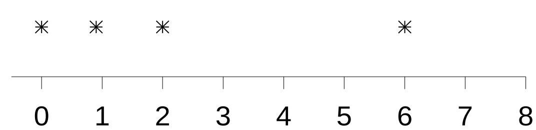
$$P(x|\mu) = \frac{1}{Z}\exp\left(-|x-\mu|\right),$$

where Z is the normalizing constant, $Z = 2$. The mean μ is not known. An example of this distribution, with $\mu = 1$, is shown below.

$P(x|\mu = 1)$



Assuming the four datapoints are



$$\{x_n\} = \{0, 0.9, 2, 6\},$$

what do these data tell us about μ? Include detailed sketches in your answer. Give a range of plausible values of μ. [8]