UNIVERSITY OF BRISTOL

January 2019 Examination Period

FACULTY OF ENGINEERING

M Level Examination for the Degree of
Bachelor of Science / Master of Engineering / Masters of Science

COMSM0021
Neural Information Processing

TIME ALLOWED:
2 hours

This paper contains *two* parts.
The first section contains *15* short questions.
Each question is worth *two marks* and all should be attempted.
The second section contains *three* long questions.
Each long question is worth *20 marks*.
The best *two* long question answers will be used for assessment.
The maximum for this paper is *70 marks*.

Other Instructions:

Calculators must have the Faculty of Engineering Seal of Approval.

TURN OVER ONLY WHEN TOLD TO START WRITING

Section A: short questions - answer all questions

Q1. What is the maximum value Shannon's entropy could have for a random variable that takes eight possible values. What is its probability distribution.

Q2. Calculate Shannon's entropy for the variable X where $p_X(a) = 1/2$, $p_X(b) = p_X(c) = 1/4$.

Q3. The infomax algorithm for recorded signals

$$\mathbf{r} = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix}$$

tries to find M that finds

$$\mathbf{x} = M\mathbf{r}$$

with x_1 and x_2 as independent as possible. It seeks to do this by maximizing $H(X_1, X_2)$; why is this useful?

Q4. In Bayesian statistics what is meant by a prior?

Q5. In the Eriksen flanker task sketch the accuracy versus reaction time for the consistent, **HHH**, and inconsistient, **HSH**, conditions. The overall scale of the reaction time is not what is being asked for, rather the shape.

Q6. What two pieces of evidence are fused in a Kalman filter?

Q7. Give two features specific to convolutional neural networks that are inspired by neuroscience.

Q8. What is the key difference between supervised and unsupervised learning in terms of their cost function?

Q9. Give an example of a gated recurrent neural networks (GRNN).

Q10. Give the value update function of temporal difference (TD) learning.

Q11. What is the role of the Dopamine neuromodulation system for reinforcement learning in the brain?

Q12. The pairwise maximum entropy model describes the probability distribution across neural population activity patterns as $P(\mathbf{x}) = \frac{1}{Z} \exp\left[\sum_i h_i x_i + \sum_{i \neq j} \frac{1}{2} J_{ij} x_i x_j\right]$ where $\mathbf{x}$ is the vector of binary neural activities and Z is a normalising constant to make sure $P(\mathbf{x})$ sums to one. What do the parameters h_i and J_{ij} represent?

Q13. Draw a diagram of the structure of a Restricted Boltzmann Machine labeling the components.

Q14. What is the neural manifold hypothesis?

Q15. Name two limitations of Principal Component Analysis for analysing neural population data.

Section B: long questions - answer two questions

Q1. This question is about sensory fusion.

(a) If we have a Markov chain

$$V \to X \to H$$

what can we say about $p_{V,H|X}(v, h|x)$? *[4 marks]*

(b) In the Ernst and Banks experiment participants are asked to assess the height of a block using visual and haptic input. If the visual estimate is v and the haptic estimate is h what is meant by the maximum likelihood estimate of the actual height x? *[4 marks]*

(c) Assumption that the visual and haptic estimates are conditionally independent, conditioned on the true value, and normally distributed about the true value with variances σ_v^2 and σ_h^2 respectively. Show the mean of $p(x|v, h)$ is

$$\mu = \frac{\sigma^2}{\sigma_v^2} v + \frac{\sigma^2}{\sigma_h^2} h$$

where

$$\frac{1}{\sigma^2} = \frac{1}{\sigma_v^2} + \frac{1}{\sigma_h^2}$$

[8 marks]

(d) In David and Burr (2004) there is a discussion of the ventriloquist effect, whereby we perceive sound as coming from a visually cued location. An example is when we watch a film and perceive a voice as coming from the person who is speaking on screen rather than the audio speaker located elsewhere in the cinema. In their experiment David and Burr present a sound and, visually, a Gaussian blob somewhere along a line on a screen. If σ_s^2 is the variance in our perception of the location of a sound source and σ_b^2 is the variance of the Gaussian blob, where would you speculate, from a Bayesian point-of-view that the participants perceive the sound as coming from? *[4 marks]*

Q2. This question is about backpropagation in the brain.

(a) In what way does the backpropagation algorithm solve the credit assignment problem? *[5 marks]*

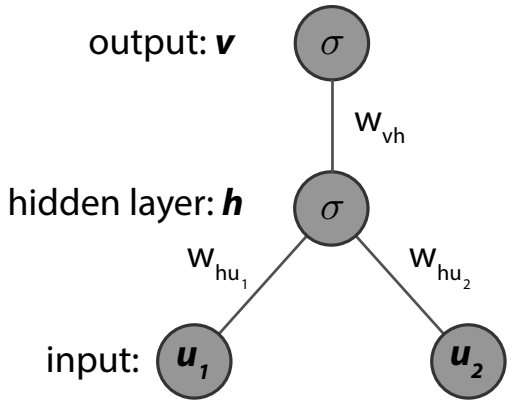

Figure 1: **Schematic of simple feedfoward neural network, with sigmoidal units** $\sigma(x)$**.**

(b) Explain three key features that have been suggested to make the backpropagation algorithm used in supervised learning biologically implausible? You should use a simple two layer neural network (with one hidden neuron h, one output neuron v, two input weights, and without biases, see Figure 1) to derive the weight updates and make a schematic of the network with the backprop to help illustrate your answer. Assume the cost function (or error) to be $E = (v - y)^2$, where y is the desired target. *[11 marks]*

(c) Which biologically implausible feature of backprop does *feedback alignment* address? And how does *feedback alignment* address it? *[4 marks]*

Q3. This question is partly about sparse coding and partly about multiunit coding.

(a) What types of features are learned by sparse coding algorithms when applied to natural images? *[4 marks]*

(b) Give the classical cost function used in sparse coding and describe its components. *[4 marks]*

(c) Imagine we have recorded the spiking patterns of a population of three neurons at eight sequential points in time. We can represent the data in a 3×8 matrix D as

$$D = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

where each row corresponds to the activity time series for one of the three neurons. Write down the time-averaged probability that a neuron is ON p_i for each of the three neurons. *[4 marks]*

(d) Use the same matrix D as in the previous question. Assume a model where each neuron is independent so that the probability of a pattern $\mathbf{x} = x_1, x_2, x_3$ is $p(\mathbf{x}) = \prod_i [p_i x_i + (1 - p_i)(1 - x_i)]$. This model can be used to compute the probability of any activity pattern, even if it was not observed in the original data. Compute the probability of the pattern $0, 1, 1$. *[4 marks]*

(e) Referring to the model in the previous question, the three neurons can jointly make $2^3 = 8$ possible binary patterns. Which of the eight patterns is/are the most probable under the independent model? *[4 marks]*

END OF PAPER