COMS10011 sample paper

THIS IS STILL A WORK IN PROGRESS but is close to its final form! This is a sample paper, it has the same style of question as the real paper, the layout is slightly different in trivial ways to the official exam layout.

Rubric

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*.
Calculators must have the Faculty of Engineering Seal of Approval.

Section A: short questions - answer all questions

1. What is the definition of Shannon's entropy for a discrete distribution?

2. If you have a finite set of n spike trains and calculate the entropy of the spike trains using the discretization method proposed by Bialek and co-workers, in general, what limit would the entropy reach as the time step is made very small?

3. Given a Markov chain $V \to X \to H$ what can we say about $p_{V,H|X}(v,h|x)$?

4. What is the cocktail party problem?

5. 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.

6. The n-armed bandit task is used in psychological studies of decision making. What is an n-armed bandit?

7. If four options in a decision task have estimated reward values r_1, r_2, r_3 and r_4, what is the soft-max probability for chosing the ith option?

8. Define the credit assignment problem in neuroscience.

9. What are the two key features of gated recurrent neural networks?

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

11. What types of features are learned by sparse coding algorithms when applied to natural images?

12. Give the classical cost function used in sparse coding.
13. Give the value update function used in Q-Learning? How does it differ from TD learning?
14. How many different joint activity patterns could a population of N binary neurons generate?
15. What is the basic idea underlying the dichotomised gaussian model for neural population data?

Section B: long questions - answer two questions

1. This question is about the Kalman filter.
   (a) [7 marks] Consider two random variables which are conditionally independent with normal distributions:
   $$\begin{aligned} p_{X|H}(x|h) &= \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-(x-h)^2/2\sigma_X^2} \\ p_{Y|H}(y|h) &= \frac{1}{\sqrt{2\pi\sigma_Y^2}} e^{-(y-h)^2/2\sigma_Y^2} \end{aligned} \tag{1}$$
   then $p_{X|H}(x|h)p_{Y|H}(y|h)$ is proportional to a normal distribution in h. What is that normal distribution?
   (b) [8 marks] Consider an object moving at constant speed v so that its position x after a time δt is
   $$x(t + \delta t) = x(t) + v\delta t + \xi$$
   where ξ is random noise drawn from $\mathcal{N}(0, \sigma_s^2)$. A sensor estimates the position of the object with noise drawn from $\mathcal{N}(0, \sigma_s^2)$. Derive the Kalman gain for optimally estimating the position of the object.
   (c) [5 marks] Explain what is meant by a forward model for motor control.
2. There are two parts to this question, the first is about information theory, the second is about statistical models.
   (a) Information theory
      i. [3 marks] Define $I(X;Y)$ and give a sufficient condition for $I(X;Y) = 0$ for non-trivial random variables X and Y?
      ii. [4 marks] Calculate the mutual information between random variables X and Y with sample spaces $\mathcal{X} = \{\mathrm{a}, \mathrm{b}, \mathrm{c}\}$ and $\mathcal{Y} = \{\alpha, \beta\}$.

	a	b	c
α	0.5	0.125	0
β	0	0.125	0.25

You can write the answer in terms of $\log 3$ and $\log 5$ if you would prefer.

iii. [3 marks] Show $I(X;Y) = H(X) - H(X|Y)$.

(b) Statistical models.

The pairwise and K-pairwise maximum entropy statistical models are often used for neural population data. The K-pairwise model gives the probability for a neural population activity pattern as

$$p(x_1, x_2, \dots, x_N) = \frac{1}{Z}\exp[\sum_i h_i x_i + \sum_{i \neq j} J_{ij} x_i x_j + V(\sum_i x_i)]$$

i. [3 marks] The equation for the standard pairwise maximum entropy model differs from the above by one term. Which term does it omit?

ii. [6 marks] How many unique parameters do each of the pairwise and K-pairwise maximum entropy models have?

iii. [1 mark] Which of the two models more accurately matches neural population data? [1 mark]

3. This question is about supervised learning in the brain.

(a) [5 marks] How may the brain implement supervised learning? [5 marks]

(b) [11 marks] Give one feature that has 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, one input weights, and no biases, see Figure 1) to derive the weight update. Assume the cost function (or error) to be $E = (v - y)^2$, where y is the desired target [11 marks].

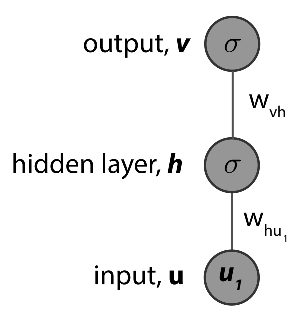

Figure 1: **Schematic of simple feedfoward neural network, with quadratic units $\sigma(x) = x^2$.**

(c) [4 marks] What features do artificial neuronal networks trained with backprop learn (e.g. when trained to discriminate objects in images)? Why do they provide a good match to the activity of neurons in the brain? [4 marks]