

ELL457/HSL622: Major Test

May 3, 2023

Maximum Marks: 26

Instructions:

- Please clearly indicate the question number, and part number if applicable, at the start of each response.
- Please read all questions carefully.
- Please ensure that your responses are to-the-point and that you write only what is asked for on the answer script you submit.
- While the exam is open-notes, all your answers must be written entirely in your own words, without any copying from anywhere.
- Please try to be clear and careful with all formal/mathematical notation, so that there is no ambiguity in the expressions/formulae you write down. Try to stick to the kind of notation used in class as far as possible.

1. Suppose I come to you with some examples of responses from ChatGPT which display a lack of understanding of figurative language: things like poetry and metaphor. I claim that this is a validation of John Searle's Chinese Room argument, showing that mere symbol-processing machines can never have true understanding.
 - (a) Is it really a validation of Searle's argument? Explain precisely why or why not. [3]
 - (b) How do you think a believer in the Turing test as an assessment of understanding/intelligence would respond to my claim? [3]
2. (a) Consider the following two sentence completion tasks:
 - i. The keys to the cabinet _____ on the table.
 - ii. These cabinets _____ all empty.What is the key difference between these two tasks in terms of the nature of the cognitive processing involved? Which of these examples could be used to argue for connectionist modelling, in the context of the Second AI Debate? What kind of connectionist model, and why? [3]
- (b) How can thinking about the nature and role of *representations* provide a way of reconciling the two sides of the Second AI Debate? Please give a focused answer: you are NOT being asked for a general description or discussion of the debate. [4]
3. Suppose I have shown a set of images to some monkeys and recorded their visual cortex responses to each image using a multielectrode array. In this data, I notice that there is a specific group of neurons which display high activity only for images of fruits and not for any of the others in my image set. How can I test whether a given CNN model might replicate this specific aspect of the brain's visual representations? Please give a precise experiment (which should be as simple as possible) to be carried out for this particular situation, and NOT a general method of assessing representational similarity. [3]
4. Using the kinds of linear threshold units discussed in class (please stick to the same notation), design an artificial neural network to compute the following Boolean function (i.e., both inputs and outputs are 0/1). Try to make your design as simple as possible, e.g., minimising the number of hidden layers.

$$f(x_1, x_2, x_3, x_4) = (x_1 \text{ OR } (\text{NOT } x_2)) \text{ OR } ((\text{NOT } x_3) \text{ OR } x_4)$$

- (a) Please draw the full neural network clearly, explicitly showing the values of all edge weights and activation thresholds. The inputs to the network should be x_1, x_2, x_3 , and x_4 , and the output should be the value of the above function. [4]
- (b) Is the above function a linear threshold function? Give a clear proof/justification, with reference to your neural network in part (a). [3]
5. Consider the Bayesian hierarchical model for semantic memory we discussed in class. Suppose I have trained that model on a corpus of documents to infer the following three topics (for each topic, the 10 most frequent words are listed):
- Topic 1: *play, stage, audience, theatre, actors, drama, performance, costumes, comedy, tragedy.*
 - Topic 2: *team, game, cricket, football, tennis, player, play, ball, field, court.*
 - Topic 3: *judge, trial, case, jury, court, accused, guilty, defendant, lawyer, justice.*

Now consider the following new document:

Last evening's performance of the classic Shakespearean tragedy Romeo and Juliet at the Kamani Theatre was very well-received by the audience. The star-cast of actors, adorned in period costumes, movingly brought to life the drama of the fates of the star-crossed lovers in the timeless play.

What will the posterior distribution over the topic to which the word *play* belongs in this document look like? (Assume that the prior distribution assigned equal probability to all 3 topics.) Explain precisely why, based on your knowledge of how the model works. [3]