COMPUTER SCIENCE TRIPOS Part IB

Wednesday 3 June 2020 1.30 to 4.30

COMPUTER SCIENCE Paper 6

*Answer **five** questions.*

*Submit the answers in five **separate** bundles, each with its own cover sheet. On each cover sheet, write the numbers of **all** attempted questions, and circle the number of the question attached.*

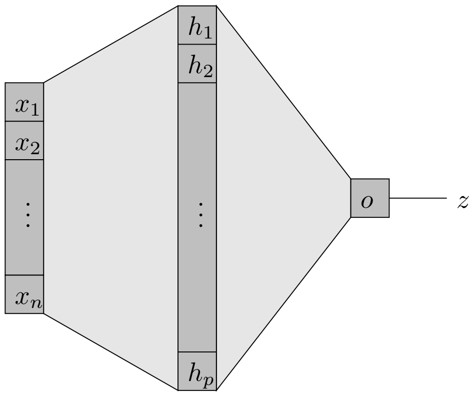
You may not start to read the questions printed on the subsequent pages of this question paper until instructed that you may do so by the Invigilator

STATIONERY REQUIREMENTS
Script paper
Blue cover sheets
Tags

SPECIAL REQUIREMENTS
Approved calculator permitted

1 Artificial Intelligence

A neural network takes input vectors $\mathbf{x} \in \mathbb{R}^n$, has a single layer of hidden nodes h_i where $i = 1, \ldots, p$, and a single output node o.



All nodes compute the function $z = \sigma(a)$ where

$$a = \sum_{i=1}^{m} w_i z_i + w_0.$$

Here z and z_i denote the inputs and outputs of the node, and each node has its own set of weights $w_0, w_1, \ldots, w_m$. Examples take the form $(\mathbf{x}, y)$ and the error the network makes for an example is $E(\mathbf{x}, y, \mathbf{w})$, where $\mathbf{w}$ is the collection of all the weights in the network.

(*a*) An example has been applied to the network and we know the quantity $\delta = \partial E / \partial a$ for the output node o. Explain how this knowledge can be used to compute the partial derivative of E with respect to the weights for some hidden node h_i. [8 marks]

(*b*) The hidden nodes are now replaced with a different type of node, computing

$$z_i = h_i(\mathbf{x}) = \phi(||\mathbf{x} - \mathbf{c}_i||^2)$$

where ϕ is some new function, the $\mathbf{c}_i \in \mathbb{R}^n$ are the parameters for the new nodes, and

$$||\mathbf{v}||^2 = \mathbf{v}^T \mathbf{v}.$$

Give a detailed derivation of a training algorithm for this network. [12 marks]

2 Artificial Intelligence

A *constraint satisfaction problem (CSP)* has four variables V_1, V_2, V_3, V_4, each with domain $\{1, 2\}$. The constraints for the problem require that given any three variables exactly one must have the value 1.

(*a*) Explain how this problem can be represented as a CSP that uses only binary constraints. Illustrate your answer by giving a graph representing the problem. [4 marks]

(*b*) Describe how *forward checking* can be used to aid the search for a solution to a CSP. Illustrate your answer by showing how it applies to the problem in Part (*a*), for assignments $V_1 = 1$ followed by $V_2 = 2$. [4 marks]

(*c*) Describe the *AC-3* algorithm for imposing consistency in a CSP. Include in your answer descriptions of an *arc*, what it means for an arc to be *consistent*, how a non-consistent arc can be made consistent, and the overall operation of the algorithm. [6 marks]

(*d*) Consider again the problem in Part (*a*). We initially have no assignments, and start by setting $V_1 = 1$. Explain in detail what happens if we attempt to adjust the domains to impose consistency. [6 marks]

3 Complexity Theory

Recall that, for languages A and B, $A \leq_P B$ denotes that A is *polynomial-time reducible* to B.

(*a*) Show that if $A \leq_P B$ and $B \leq_P C$, then $A \leq_P C$. [3 marks]

(*b*) Show that if A is NP-complete and $A \leq_P B$ and $B \leq_P A$, then B is also NP-complete. [3 marks]

For any $k \in \mathbb{N}$, let k-EC (*exact cover by k sets*) denote the language coding the following decision problem.

> Given a set X and a collection S of k-element subsets of X, determine if there exists a set $C \subseteq S$ such that each $x \in X$ appears in exactly one element of C.

(*c*) Show that, for each fixed k, k-EC $\leq_P$ $(k+1)$-EC. [5 marks]

For each of the following statements, determine whether the it is true, false or unknown and give full justification for your answer. If the answer is unknown, the justification may be that it is equivalent to some well-known open problem. You may assume standard facts about the complexity of matching and covering problems covered in the lectures, as long as you state them clearly.

(*d*) 2-EC $\leq_P$ 4-EC [3 marks]

(*e*) 4-EC $\leq_P$ 3-EC [3 marks]

(*f*) 4-EC $\leq_P$ 2-EC [3 marks]

4 Complexity Theory

Let $f : \mathbb{N} \to \mathbb{N}$ be a *constructible* function.

(*a*) Explain why $\text{NT}{\small\text{IME}}(f) \subseteq \text{S}{\small\text{PACE}}(f)$ [4 marks]

(*b*) Explain why any language $L \in \text{NS}{\small\text{PACE}}(f)$ is also in $\text{T}{\small\text{IME}}(c^{(f(n)+\log n)})$ for some constant c. [6 marks]

(*c*) A deterministic machine is a special case of a non-deterministic machine. Expain what this and the results above tell us about the inclusions among the following complexity classes:

L, NL, P, NP, PSpace, NPspace.

[5 marks]

(*d*) Savitch proved that there is an algorithm for the graph reachability problem that uses $O((\log n)^2)$ space. What further inclusions can you derive among the above complexity classes using this fact? Explain your answer. [5 marks]

5 Computation Theory

(*a*) Explain what it means for a partial function $f : \mathbb{N}^n \rightharpoonup \mathbb{N}$ to be register machine *computable* and for a set of numbers $S \subseteq \mathbb{N}$ to be register machine *decidable*. [5 marks]

(*b*) A set of numbers $S \subseteq \mathbb{N}$ is register machine *enumerable* if either S is empty, or $S = \{f(n) \mid n \in \mathbb{N}\}$ for some total function $f : \mathbb{N} \to \mathbb{N}$ which is register machine computable.

(*i*) Show that if S is register machine decidable, then it is register machine enumerable. [5 marks]

(*ii*) Show that if both S and its complement $\overline{S} \triangleq \{n \in \mathbb{N} \mid n \notin S\}$ are register machine enumerable, then S is register machine decidable. [5 marks]

(*iii*) Give an example of a set of numbers that is register machine enumerable, but not register machine decidable. (Any standard results about computable functions that you use should be carefully stated.) [5 marks]

6 Computation Theory

(*a*) Define the Church numerals for zero ($\underline{0}$), one ($\underline{1}$) and for an arbitrary natural number ($\underline{n}$). [2 marks]

(*b*) Define encodings of Booleans as λ-terms (True, False and If). [1 mark]

(*c*) Explain what it means for a λ term to *represent* a number-valued partial function of n numerical arguments; do the same for one returning Boolean instead of numerical results. [3 marks]

(*d*) Give λ-terms that represent the following functions:

(*i*) successor (Succ) [1 mark]

(*ii*) test for zero (Eq_0) [1 mark]

(*e*) Define encodings of pairing and projections (Pair, Fst and Snd). [2 marks]

(*f*) What function $\mathbb{N} \to \mathbb{N}$ is represented by the following λ-term? Carefully justify your answer.

$$\lambda x.\, \mathsf{Snd}(x(\lambda y.\, \mathsf{Pair}(\mathsf{Succ}(\mathsf{Fst}\, y))(\mathsf{Fst}\, y))(\mathsf{Pair}\, \underline{0}\, \underline{0}))$$

[6 marks]

(*g*) Give with justification a λ-term that represents the function mapping each pair of numbers (m, n) to True if $m \leq n$ and to False otherwise. [Hint: use the λ-terms from parts (*d*)(*ii*) and (*f*).] [4 marks]

7 Foundations of Data Science

Consider the probability model

$$
\begin{array}{ccccccccc}
X_0 & \to & X_1 & \to & X_2 & \to & X_3 & \to \cdots \\
 & & \downarrow & & \downarrow & & \downarrow & \\
 & & Y_1 & & Y_2 & & Y_3 &
\end{array}
$$

where $(X_0, X_1, \dots)$ is a Markov chain on state space $\{0, 1\}$ with transition probabilities $P_{01} = p$, $P_{10} = q$; and where each Y_i is normally distributed with mean X_i and variance σ^2.

We are given a sequence of observations $(y_1, y_2, \dots, y_n)$, and we wish to make an inference about the unobserved values $(X_1, X_2, \dots, X_n)$. We will take $0 < p < 1$, $0 < q < 1$, and $\sigma > 0$ to be known, and we will assume that X_0 is sampled from the Markov chain's stationary distribution.

(*a*) Write out the transition matrix for the Markov chain $(X_0, X_1, \dots)$. Calculate its stationary distribution. [4 marks]

(*b*) Writing $\vec{X}$ for $(X_0, X_1, \dots, X_n)$, and writing $\vec{Y}$ for $(Y_1, \dots, Y_n)$, and similarly $\vec{x}$ and $\vec{y}$, find expressions for

$$\mathbb{P}(\vec{X} = \vec{x}) \quad \text{and for} \quad \mathbb{P}(\vec{Y} = \vec{y} \mid \vec{X} = \vec{x}).$$

[4 marks]

(*c*) Give pseudocode for a function `rx(n)` that generates a random $\vec{X}$. Give pseudocode to generate a weighted sample from the posterior distribution of $\vec{X}$ conditional on the observed data $\vec{Y} = \vec{y}$. [8 marks]

(*d*) Let $Z = n^{-1} \sum_{i=1}^{n} X_i$. Give pseudocode to find a 95% confidence interval for Z, conditional on the observed data $\vec{Y} = \vec{y}$. [4 marks]

8 Foundations of Data Science

This table shows a summary of temperature readings from the Cambridge weather station, comparing June, July, and August in the 1970s to the 2010s. It shows the number of months in which the average maximum daily temperature was low (< 15.5°C), high (> 18°C), or medium. We wish to establish whether there is a signficant difference between the two rows.

	low	med	high
1970s	10	18	2
2010s	5	14	11

Suppose that the readings are independent from month to month. Let $p_{d,k}$ be the probability that a month's reading falls into bin $k \in \{\text{low}, \text{med}, \text{high}\}$, in decade $d \in \{1970\text{s}, 2010\text{s}\}$. The $p_{d,k}$ are unknown parameters.

(*a*) Give expressions for the maximum likelihood estimates $\hat{p}_{d,k}$. In your answer, you should state what is being maximized, over what variables. [3 marks]

(*b*) Let the null hypothesis H_0 be that the probabilities are the same in the 1970s as in the 2010s; call these common probabilities q_k. Give expressions for the maximum likelihood estimates $\hat{q}_k$ under H_0. [2 marks]

(*c*) We wish to test H_0, using the test statistic

$$t = \sum_{d,k} \frac{(\hat{p}_{d,k} - \hat{q}_k)^2}{\hat{q}_k}.$$

(*i*) Explain briefly what is meant by *parametric resampling*. Explain how to compute the distribution we'd expect to see for t, under H_0. Give pseudocode. [6 marks]

(*ii*) Explain what is meant by a one-sided test versus a two-sided test. Which should we use in this case? [3 marks]

(*iii*) Give pseudocode to compute the p-value of this test. [3 marks]

(*d*) What are some advantages and disadvantages of this count-based test, compared to a test based on linear regression? [3 marks]

9 Logic and Proof

(*a*) In the context of automatic theorem proving, provide one-sentence definitions of each of the following concepts: *satisfiable*, *sound*, *complete*. You may take as given the definitions of all underlying concepts. [3 marks]

(*b*) Mordred has written a resolution theorem prover, but there are bugs in his code. Very rarely, one of the following errors occurs: a literal is deleted from a clause; an entire clause is deleted; the "occurs check" of unification is not performed. Briefly describe, with justification, the consequences of each type of error. [3 marks]

(*c*) For each of the following sets of clauses, either derive the empty clause or demonstrate that the set is satisfiable by exhibiting a model. Below, a and b are constants, while x, y and z are variables.

(*i*)

$$\begin{array}{rl} \{R(a)\} & \{\neg R(x), \neg Q(f(x)), \neg R(a)\} \\ \{Q(z), P(z)\} & \{\neg P(y), \neg R(y)\} \end{array}$$

[7 marks]

(*ii*)

$$\begin{array}{rl} \{R(a), R(b)\} & \{\neg R(x), \neg Q(f(x)), \neg R(y)\} \\ \{Q(x), \neg P(y)\} & \{Q(z), P(z)\} \end{array}$$

[7 marks]

10 Logic and Proof

(*a*) From n distinct propositional letters, each of which may be negated or not, 2^n distinct clauses can be created. Present a satisfying interpretation of this set of 2^n clauses or demonstrate that none exists. [3 marks]

(*b*) Sketch the operation of the DPLL algorithm when provided with the set of clauses described above, including an estimate of its time complexity as a function of n. [4 marks]

(*c*) For each of the following formulas, present either a proof in the sequent calculus, or a falsifying interpretation. The modal logic is S4.

(*i*) $\Box(P \lor Q) \to (\Box\Diamond\neg P \to \Diamond\Box Q)$ [5 marks]

(*ii*) $\exists x\, P(f(x)) \land \forall x\, [P(x) \to Q(g(x))] \to \exists y\, Q(y)$ [4 marks]

(*iii*) $\exists x\, (P(x) \to Q(x)) \to [\exists x\, P(x) \to \exists x\, Q(x)]$ [4 marks]

END OF PAPER