CST1
COMPUTER SCIENCE TRIPOS Part IB

Wednesday 8 June 2022 11:00 to 14:00 BST

COMPUTER SCIENCE Paper 6

Answer ***five*** *questions.*

Submit each question answer in a ***separate*** *PDF. As the file name, use your candidate number, paper and question number (e.g.,* `1234A-p6-q6.pdf`*). Also write your candidate number, paper and question number at the start of each PDF.*

You must follow the official form and conduct instructions for this online examination

1 Artificial Intelligence

You are solving a planning problem using the GraphPlan algorithm. The problem has the following actions:

$\neg A(S), H(M)$	$\neg A(B)$	$\neg H(M), A(B)$	$\neg H(C), H(M), A(S)$
$\boxed{a_1(S)}$	$\boxed{a_1(B)}$	$\boxed{a_2(M)}$	$\boxed{a_2(S)}$
$A(S)$	$A(B)$	$H(M)$	$H(C), \neg H(M)$

The start state is $\neg H(M), \neg H(C), \neg A(S), \neg A(B)$ and the goal is $H(C)$.

(*a*) Draw the state levels S_i and action levels A_i of the planning graph up to and including S_2. Do not add any mutexes at this stage. [3 marks]

(*b*) Explain why it is not possible at this stage to attempt to extract a plan. [1 mark]

(*c*) What is the smallest i for which it might make sense to try to extract a plan, starting from S_i. Explain your answer. [2 marks]

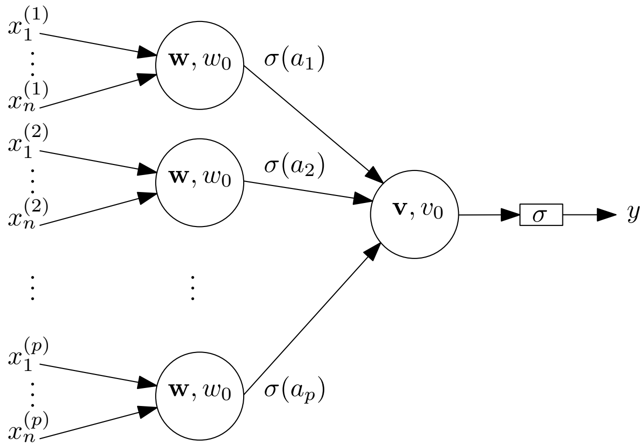
(*d*) For the value of i identified in Part (*c*), draw levels S_{i-1}, A_{i-1} and S_i of the planning graph. [4 marks]

(*e*) On the diagram you have produced for Part (*d*), mark four mutexes, each of which arises for a different reason. In each case explain what kind of mutex you have included. [6 marks]

(*f*) Give a general description of how the extraction of a plan from a planning graph can be addressed as a heuristic search problem. [4 marks]

2 Artificial Intelligence

This question addresses a variation on the usual *multilayer perceptron*.



The input vector is divided into p groups each with n elements. Let $x_i^{(j)}$ be the ith element in the jth group and let $\mathbf{x}^{(j)} = \begin{pmatrix} x_1^{(j)} & \cdots & x_n^{(j)} \end{pmatrix}^T$. There are p nodes in the hidden layer, which share a single weight vector $\mathbf{w}$ and bias w_0. Thus the kth hidden node computes $\sigma(a_k)$ where σ is an activation function and

$$a_k = \mathbf{w}^T \mathbf{x}^{(k)} + w_0.$$

Let $\mathbf{a} = \begin{pmatrix} a_1 & \cdots & a_p \end{pmatrix}^T$. The output node then combines the hidden nodes in the usual way using weights $\mathbf{v}$ and v_0 to produce $y = \sigma(a)$ where

$$a = \sum_{i=1}^{p} v_i \sigma(a_i) + v_0.$$

Collecting all the parameters of the network into a single vector $\boldsymbol{\theta}$, the error for a single labelled example is $E(\boldsymbol{\theta})$.

(*a*) Show that the value of $\delta = \partial E(\boldsymbol{\theta})/\partial a$ is

$$\delta = \sigma'(a) \frac{\partial E(\boldsymbol{\theta})}{\partial y}.$$

[3 marks]

(*b*) Find expressions for the partial derivatives of $E(\boldsymbol{\theta})$ with respect to the parameters of the single output node. [5 marks]

(*c*) Show that the partial derivatives $\delta_i = \partial E(\boldsymbol{\theta})/\partial a_i$ for the hidden nodes are

$$\delta_i = \delta v_i \sigma'(a_i).$$

[5 marks]

(*d*) Find expressions for the partial derivatives of $E(\boldsymbol{\theta})$ with respect to the parameters of the hidden nodes. [7 marks]

3 Complexity Theory

(*a*) If A and B are decision problems, we write $A \leq_L B$ to denote that A is reducible to B by means of a *logarithmic-space* reduction. Give a precise definition of such a reduction. [2 marks]

(*b*) For decision problems A, B and C, show that if $A \leq_L B$ and $B \leq_L C$, we have $A \leq_L C$. [5 marks]

(*c*) For each of the four complexity classes P, NP, NL and $\mathsf{co\text{-}NP}$, give an example of a problem that is *complete* for the complexity class under logarithmic-space reductions. You do not need to prove the completeness. [4 marks]

(*d*) For each pair of problems A and B from your answers to part (*c*) above, state whether or not $A \leq_L B$, or if this is unknown. Where it is unknown, state any consequences about the inclusion of complexity classes that would follow from $A \leq_L B$. [9 marks]

4 Complexity Theory

For the purpose of this question, a graph $G = (V, E)$ is a set V of vertices along with a set E of edges where each edge is a set of two *distinct* vertices. That is, we consider undirected graphs without self-loops or multiple edges.

Given two graphs $G = (V, E)$ and $H = (U, F)$, a *homomorphism* from G to H is a function $h : V \to U$ such that whenever $\{v_1, v_2\}$ is in E, $\{h(v_1), h(v_2)\}$ is in F. We write HOM for the decision problem consisting of all pairs of graphs (G, H) such that there is a homomorphism from G to H.

Recall that a graph $G = (V, E)$ is k-colourable (for a positive integer k) if there is a function $\chi : V \to \{1, \ldots, k\}$ such that whenever $\{u, v\}$ is in E, $\chi(u) \neq \chi(v)$.

(a) Explain why the decision problem HOM is in NP. [4 marks]

(b) Let K_3 denote the graph with three vertices a, b, c and the three edges $\{a, b\}, \{b, c\}$ and $\{a, c\}$. Show that for any graph G, there is a homomorphism from G to K_3 if, and only if, G is 3-colourable. [6 marks]

(c) What can you conclude from the above about the complexity of the problem HOM? [5 marks]

(d) Let K_2 denote the graph consisting of two vertices a and b and the single edge $\{a, b\}$. What is the complexity of the decision problem consisting of all graphs G for which there is a homomorphism from G to K_2? [5 marks]

5 Computation Theory

Given a partial function $f : \mathbb{N} \rightharpoonup \mathbb{N}$, let $D(f)$ denote the set of natural numbers at which f is defined: $D(f) = \{x \in \mathbb{N} \mid f(x)\downarrow\}$; and let $I(f)$ be the set of natural numbers that are values of f where it is defined: $I(f) = \{y \in \mathbb{N} \mid \text{for some } x,\ f(x) = y\}$.

Prove or disprove the following statements, clearly stating any results about register machine computable functions and partial recursive functions that you use.

(*a*) Every subset $S \subseteq \mathbb{N}$ is equal to $I(f)$ for some register machine computable partial function f. [4 marks]

(*b*) If f is register machine computable, then $I(f)$ is equal to $D(g)$ for some partial recursive function g. [7 marks]

(*c*) If f is register machine computable, then $D(f)$ is equal to $I(g)$ for some *total* recursive function g. [4 marks]

(*d*) If g is a partial recursive function, then $D(g)$ is equal to $I(f)$ for some register machine computable partial function f. [5 marks]

6 Computation Theory

(*a*) Explain why the Church-Rosser Theorem implies that any λ-term that is β-convertible ($=_\beta$) to a term in β-normal form is in fact β-reducible ($\twoheadrightarrow$) to one in β-normal form. [2 marks]

(*b*) Let *Bnf* denote the set of λ-terms that have a β-normal form. Give with justification an example of two closed λ-terms I and Ω with $\mathsf{I} \in \mathit{Bnf}$ and $\Omega \notin \mathit{Bnf}$. [3 marks]

(*c*) Suppose that $\#$ is a bijection between the set of all λ-terms and the set $\mathbb{N}$ of all natural numbers and that there are recursive functions $\alpha : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ and $\nu : \mathbb{N} \to \mathbb{N}$ satisfying for all λ-terms M, N and numbers n that

$$\alpha(\#(M), \#(N)) = \#(M\,N) \tag{1}$$
$$\nu(n) = \#(\underline{n}) \tag{2}$$

where the λ-term $\underline{n}$ is the n^{th} Church numeral. Writing $\ulcorner M \urcorner$ for $\underline{\#(M)}$, show that there are closed λ-terms App and Num satisfying for all λ-terms M, N that

$$\mathsf{App}\,\ulcorner M \urcorner \ulcorner N \urcorner =_\beta \ulcorner M\,N \urcorner \tag{3}$$
$$\mathsf{Num}\,\ulcorner N \urcorner =_\beta \ulcorner\ulcorner N \urcorner\urcorner \tag{4}$$

(Any general properties of partial recursive functions with respect to λ-calculus you use should be carefully stated, but need not be proved.) [6 marks]

(*d*) Consider the following property of a closed λ-term F (where $\ulcorner M \urcorner$ is as in part (*c*)):

$$\text{for all } \lambda\text{-terms } M, \quad \begin{cases} F\,\ulcorner M \urcorner =_\beta \underline{0} & \text{if } M \in \mathit{Bnf} \\ F\,\ulcorner M \urcorner =_\beta \underline{1} & \text{if } M \notin \mathit{Bnf} \end{cases} \tag{7}$$

Let $H = \lambda h.\mathsf{P}\,\Omega\,\mathsf{I}\,(F\,(\mathsf{App}\,h\,(\mathsf{Num}\,h)))$ where $\mathsf{P} = \lambda x\,y\,f.f\,(\lambda z.y)\,x$ and Ω and I are as in part (*b*). By considering whether or not $H\,\ulcorner H \urcorner$ is in *Bnf*, deduce that there can be no λ-term F satisfying (7). [6 marks]

(*e*) Deduce from part (*d*) that $\{\#(M) \mid M \in \mathit{Bnf}\}$ is an undecidable set of numbers. [3 marks]

7 Data Science

We are given a numerical dataset $\{x_1, x_2, \ldots, x_n\}$. We wish to estimate the 99$^{\text{th}}$ percentile, and to find a confidence interval for it. Here are three approaches:

(*a*) We may decide to model the datapoints as independent samples from the Pareto$(1, \alpha)$ distribution. Then, the 99$^{\text{th}}$ percentile is the value q such that $\mathbb{P}\big(\text{Pareto}(1, \alpha) \leq q\big) = 0.99$.

(*i*) Find the maximum likelihood estimator for α. [3 marks]

(*ii*) Find q as a function of α. [2 marks]

(*iii*) Explain how to use parametric resampling to find a confidence interval for q. Give pseudocode. [4 marks]

(*b*) We may decide to estimate the 99$^{\text{th}}$ percentile by simply sorting the dataset and reading off the value in position $\text{int}(0.99n)$.

Explain how to use nonparametric resampling to find a confidence interval for it. Give pseudocode. Under what circumstances would you expect the result to be unreliable? [6 marks]

(*c*) We may decide to use computational Bayesian methods to find the confidence interval. Explain how, stating your model precisely. Give pseudocode. [5 marks]

Hint. If $X \sim \text{Pareto}(1, \alpha)$ *then it has cumulative distribution function*

$$\mathbb{P}(X \leq x) = \begin{cases} 1 - x^{-\alpha} & \text{if } x \geq 1 \\ 0 & \text{if } x < 1. \end{cases}$$

8 Data Science

Let $(E_0, E_1, \dots)$ be a Markov chain generated by

$$E_{t+1} = \lambda E_t + \text{Normal}\big(0, (1-\lambda^2)\sigma^2\big)$$

where $0 \le \lambda < 1$.

(*a*) What is meant by "stationary distribution"? Show that $\text{Normal}(0, \sigma^2)$ is a stationary distribution for this Markov chain. [3 marks]

(*b*) What is meant by "memorylessness"? Give an expression for the log likelihood of a sequence of values $(e_1, \dots, e_n)$, given the value for E_0. [4 marks]

Klaus Hasselmann, who won the 2021 Nobel Prize, studied climate models in which short-term random fluctuations can have longer-term effects. Suppose we're given a dataset $(y_0, y_1, \dots, y_n)$ of temperatures at timepoints $t = 0, 1, \dots, n$, and we use a Hasselmann-style model,

$$Y_t = \alpha + \beta \sin(2\pi\omega t) + \gamma t + E_t$$

where $(E_0, E_1, \dots)$ is a Markov chain as described above, and α, β, and γ are unknown parameters to be estimated.

(*c*) Give an expression for the log likelihood of $(y_1, \dots, y_n)$, given the value for Y_0. [*Hint:* First find the distribution of Y_{t+1} given Y_t.] [6 marks]

(*d*) What is meant by a "linear model"? Write out a linear model that can be used to estimate the unknown parameters α, β, and γ (treating all other parameters as known). Identify the feature vectors. [7 marks]

9 Logic and Proof

(a) You have to write a program to classify given propositional formulas as unsatisfiable, valid or neither. You have the option of using either BDDs or DPLL for this. For each option, sketch how to carry out the classification and comment briefly on its advantages (e.g. from the standpoint of performance or space) compared with the alternative. [4 marks]

(b) It is possible to modify DPLL such that, whenever the empty clause is encountered within the search, a subset of the current variable assignments are identified that contradict the original set of clauses. These assignments are then used to create a new clause. For example, if P true, Q false, S false contradict the original clauses, then the new clause $\{\neg P, Q, S\}$ is added to the others.

(i) In what sense is adding this clause logically correct? [2 marks]

(ii) How is this new clause likely to affect the operation of DPLL? [2 marks]

(c) For each of the following formulas, either exhibit a formal proof in the free-variable tableau calculus, or exhibit a falsifying interpretation (a is a constant):

(i) $[\neg P(a) \wedge \forall x\, (\neg P(f(x)) \to P(x))] \to \exists x (P(x) \wedge \neg P(f(x)))$ [6 marks]

(ii) $(\forall x\, P(x)) \wedge (\forall x\, Q(x)) \vee \exists x (P(x) \to \neg Q(x))$ [6 marks]

10 Logic and Proof

(*a*) List three significant differences between the DPLL method and resolution. [3 marks]

(*b*) To what extent is pure literal elimination applicable to (i) resolution theorem proving and (ii) Prolog? [3 marks]

(*c*) For each of the following sets of clauses, either exhibit a model or show that none exists using resolution. Note: a and b are constants, while x and y are variables.

(*i*)

$$\{P(a)\} \qquad \{\neg P(x), M(x)\} \qquad \{\neg P(x), L(x)\} \qquad \{\neg Q(x), \neg P(x)\}$$
$$\{\neg L(x), \neg M(x), \neg R(x)\} \qquad \{\neg P(x), Q(x), P(b)\} \qquad \{\neg P(x), Q(x), R(b)\}$$

[7 marks]

(*ii*)

$$\{\neg P(x), Q(x,x)\} \qquad \{\neg Q(x,y), \neg Q(y,x), R(x,y)\}$$
$$\{\neg R(x,y), \neg R(a,x)\} \qquad \{P(a), P(b)\}$$

[7 marks]

END OF PAPER