Australian National University

THE AUSTRALIAN NATIONAL UNIVERSITY

First Semester Model Examination 2016

ARTIFICIAL INTELLIGENCE

(COMP3620/COMP6320)

Writing Period: 3 hours duration
Study Period: 15 minutes duration
Permitted Materials: None
You must attempt to answer all questions

Student Number

PLEASE READ THE FOLLOWING INSTRUCTIONS CAREFULLY.

- Students are asked to check that this examination paper contains all 20 pages.
- Questions are of unequal value. The value of each question is shown in square brackets. Questions that are decomposed into parts show the number of marks given to each part within square brackets.
- Answer each question using the space provided. If you run out of room, then use the space provided at the end of this paper. Scrap paper will be provided.
- This examination paper is **CONFIDENTIAL** and is not to be taken from the examination room.

Official use only:

1.1	1.2	1.	2.1	2.2	2.3	2.		Total
3.1	3.2	3.3	3.	4.1	4.2	4.3	4.	

1 **Search** [25 marks]

1.1 Questions [8/25]

a. [2/25] In which situations is breadth-first search guaranteed to find a **good** solution **quickly**?

Write your answer to Question 1.1(a) here.

Shallow/short solution,
also they branching factor not too many.

b. Let h_1 be an admissible heuristic for the set of goal states {G1} and h_2 an admissible heuristic for the set of goal states {G2}. Provide the best admissible heuristics that can be constructed for the following sets of goal states: {G1,G2} and {G1,G2,G3}.

i. [2/25] {**G1,G2**}

Write your answer to Question 1.1(b.i) here.

$\min(h_1, h_2)$

ii. [2/25] {**G1,G2,G3**}

$\frac{2}{3}\min(h_1, h_2)$

Write your answer to Question 1.1(b.ii) here.

c. [2/25] In the example of Figure 1, what is the rational action for the MAX player? Justify.

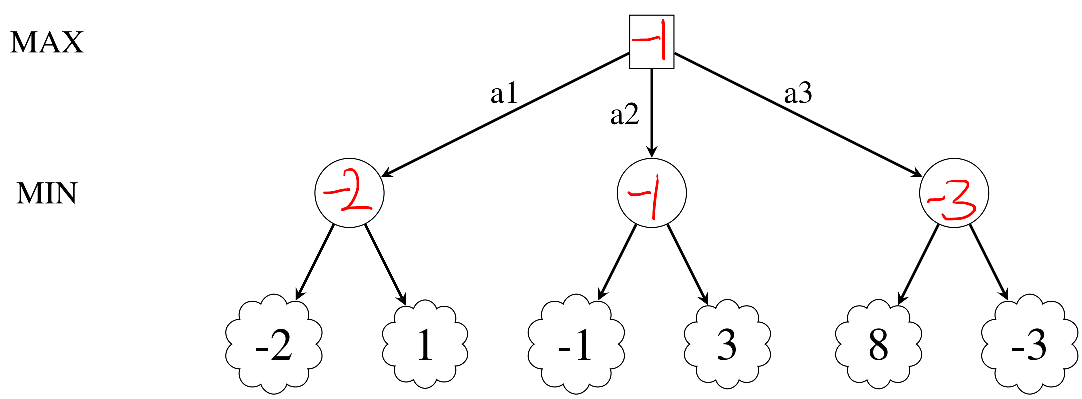


Figure 1: Game Tree

Write your answer to Question 1.1(c) here.

1.2 Search Strategies [10/25]

Consider the search problem of Figure 2 where S is the starting state and G_1, G_2 are the goal state. Arcs are labelled with the cost of traversing them and the estimated cost to a goal given by the heuristic is reported inside nodes.

B(9)
A(6)
G1(9)
B(9) G1(9) C(6) D(6)
G1(6) D(6)
G2(6)

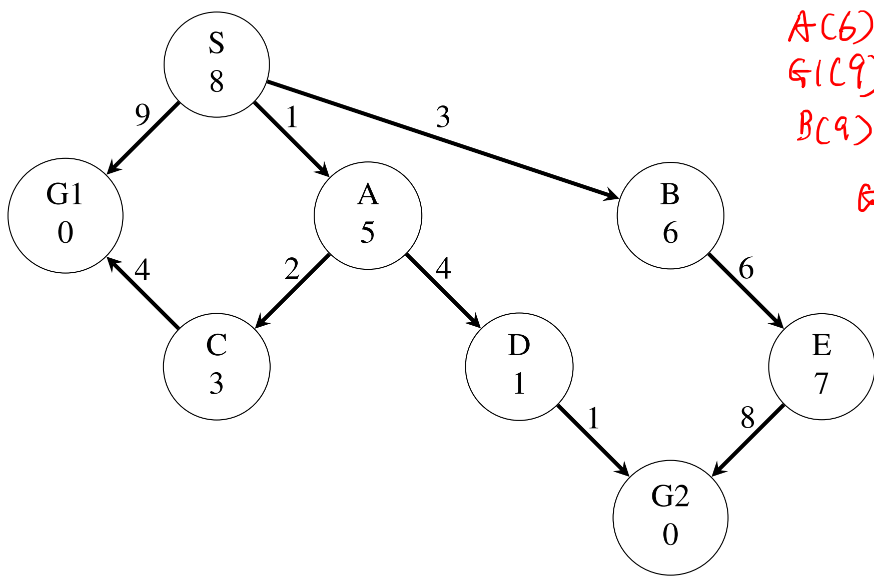


Figure 2: Search Problem

a. Compute the goal reached by applying the following search strategies: breadth-first search, iterative deepening, uniform-cost search, A*.

S
A B G1

i. [2/25] **breadth-first search**: G1

S, S
A B G1

ii. [2/25] **iterative deepening**: G1

iii. [2/25] **uniform-cost search**: G1

iv. [2/25] **A***: G1

b. Write the content of the fringe at the end of uniform-cost and A*.

i. [1/25] **uniform-cost search**:

ii. [1/25] **A***:

1.3 Stochastic Games [7/25]

a. [3/25] Apply the MINIMAX algorithm for stochastic games to the game tree of Figure 3 and write the utility value associated with each node.

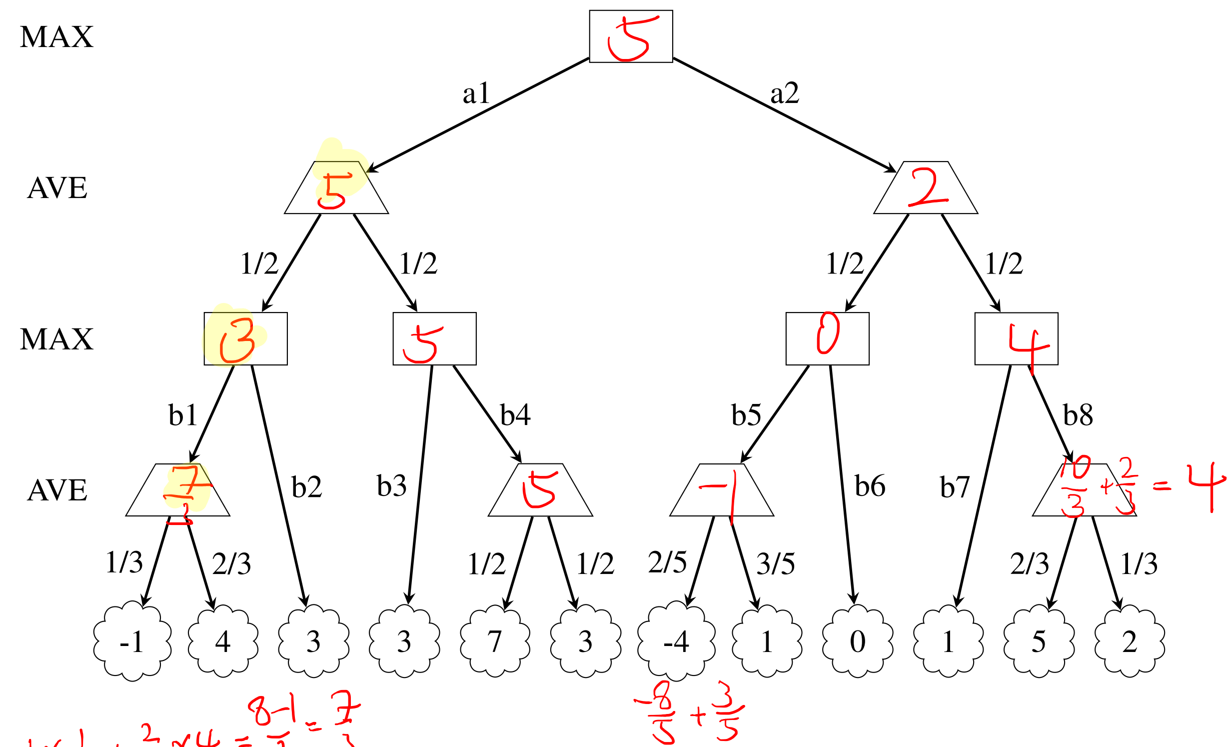


Figure 3: Game Tree

b. [2/25] What is the rational move for MAX: a1 or a2? Why?

Write your answer to Question 1.3(a) here.

c. [2/25] You are allowed to modify one of the two negative values. Which one should you change to switch the optimal move for MAX? What is the minimal integer value that will guarantee this switch?

Write your answer to Question 1.3(b) here.

change let $\frac{7}{3}$ be > 3

$-1 \Rightarrow x$

$\frac{1}{3}x + \frac{2}{3} \times 4 > 3$

$\frac{1}{3}x > 3 - \frac{8}{3} = \frac{1}{3}$

$x > 1$

$\boxed{x = 2}$

2 Knowledge Representation & Reasoning [25 Marks]

2.1 Quiz [8 marks]

a. [1/25] Which of the following

   i. $\neg\neg x \wedge \wedge y$ ✗

   ii. $\neg\neg x \rightarrow (\neg(y \vee z))$ ✓

   iii. $\neg x \wedge \neg y \vee z$ ✓

   iv. $\exists x\, [\neg x \wedge \neg y \vee z]$ ✗

   are *well–formed* propositional logic formulas?

   Write your answer to Question 2.1(a) here.

b. [3/25] Write the formula $(\neg x \wedge \neg w) \vee (y \wedge z)$ in CNF. Detail any intermediate steps.

   Write your answer to Question 2.1(b) here.

   $((\neg x \wedge \neg w) \vee y) \wedge ((\neg x \wedge \neg w) \vee z)$

   $(\neg x \vee y) \wedge (\neg w \vee y) \wedge (\neg x \vee z) \wedge (\neg w \vee z)$

c. [1/25] Let Σ be a signature with *constants* $\{Lassie, Garfield, Odie, Lasagna, Tofu\}$, *predicate symbols* $\{Dog(.), Cat(.), Eats(.,.), Chases(.,.), Equals(.,.)\}$, and *function* symbols $\{FoodOf(.)\}$. Which of the following

   i. $\forall x\, [Dog(x) \rightarrow Equals(FoodOf(x), Tofu))]$ ✓

   ii. $\exists y \forall x\, [Cat(y) \wedge Dog(x) \rightarrow Equals(FoodOf(y), Food(x))]$ ✓

   iii. $\exists x\, [FoodOf(Garfield) \wedge Dog(x)]$

   iv. $\forall y \forall x\, [Eats(Dog(y), Cat(x)]$ ✗ no function

   are *well–formed* Predicate logic formulas?

   Write your answer to Question 2.1(b) here.

d. [1/25] Given the signature Σ given in the previous question, are there any *free variable* in the following formula? If any, name them.

$$\exists x[Dog(x) \wedge Eats(x, Tofu) \wedge \forall y\,(Cat(y) \rightarrow Chases(y, z)]$$

Write your answer to Question 2.1(b) here.

z

e. [2/25] Consider the following formula, that uses the signature Σ defined above

$$\forall x \exists y \exists z[\neg Cat(x) \vee (Chases(x, y) \wedge Eats(x, FoodOf(z))]$$

Is the formula above in *Skolem Normal Form*? If not, describe its transformation into Skolem Normal Form and write the resulting transformed formula.

Write your answer to Question 2.1(b) here.

No

$\exists \neg Cat(x) \vee (Chases(x, b) \wedge Eats(x, FoodOf(c))]$

2.2 Coloring Graphs [7 Marks]

Consider the graph $G = (V, E)$ with $V = \{A, B, C\}$ and $E = \{(A, B),\ (A, C)\}$. We need to paint each vertex with a given colour - red, green or blue - keeping in mind that no two adjacent vertices can be painted with the same colour.

a. [4/25] Model this problem as a set of propositional logic formulas whose models correspond to the allowable ways to paint the vertices. Explain the meaning of each relevant formula and symbol used.

Write your answer to Question 2.2(a) here and on the next page.

b. [3/25] Then verify, with the resolution algorithm, whether it is true that when B is painted red and C is painted red, then A cannot be painted red.

Write your answer to Question 2.2(b) here.

2.3 Constraint Networks [10 marks]

a. [1/25] Find a constraint network $\gamma = (V, D, C)$ such that AC-3 reduces the domains of every variable to one single value. Describe formally V, D and C, and justify your answer.

Write your answer to Question 2.3 here.

b. [4/25] Describe he execution of AC-3 on γ by providing, for each step, the content of M, the variable pair selected, and the change, if any, in the domain of each variable.

Write your answer to Question 2.3(b) here.

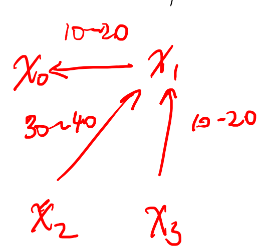
c. [5/25] Consider the following temporal constraint network $\gamma = (V, C)$ where $V = \{x_0, x_1, x_2, x_3\}$ with constraints C

i. $10 \leq x_1 - x_0 \leq 20$

ii. $30 \leq x_2 - x_1 \leq 40$

iii. $10 \leq x_3 - x_1 \leq 20$

Compute the *distance graph* G_d for γ and determine whether or not γ is *consistent*.



Write your answer to Question 2.3(b) here.

3 **Planning** [25 marks]

3.1 **Planning Questionnaire** [9 marks]

In the following questionnaire, select the correct completion of the sentence.

1. Plan-space planning is aimed at producing partial ordered plans $\langle \pi, \prec \rangle$ such that:

   a. *at least* one of the linearizations induced by $\prec$ is a valid plan

   b. *all* the linearizations induced by $\prec$ are valid plan

   c. only necessary orderings are present in $\prec$

2. Given two STRIPS actions a and b where $pre(a)$ is the precondition of a and $del(b)$ the delete list of action b, and let us consider the case where $pre(a) \cap del(b) \neq \emptyset$. We say that the two actions are in a mutex relationship because:

   a. They have inconsistent effects

   b. Action b deletes at least one of the a's preconditions

   c. Action b deletes all the preconditions of a.

3. In SAT-based planning, the explanatory frame axiom is intended to constrain the propositional theory to consider valid only the models where:

   a. actions have their preconditions satisfied in the state in which the action is applied

   b. actions have consistent effects

   c. all the fluents not affected by the actions remain unchanged

Mark your answers to Question 3.1 above.

3.2 **Modelling** [10 marks]

Consider a variation of the famous *blocks-world* domain. In this variation, actions are not the common *stack*, *unstack*, *put-down* and *pick-up*, but you have just one action encoding the swapping of two blocks. Blocks can be swapped under the conditions that *the two blocks are both clear*: this means that there is no other block on top of them. This domain is quite different w.r.t. traditional blocks-world as only the *surface* of the pile of blocks can be modified. Encode this planning problem in STRIPS or PDDL, and in particular identify:

1. A set of predicates/propositions; you can the same set of propositions used for *blocks-world*.

2. The set of actions, and their structure in terms of preconditions, positive and negative effects

3. The initial state and goal as for Figure 4.

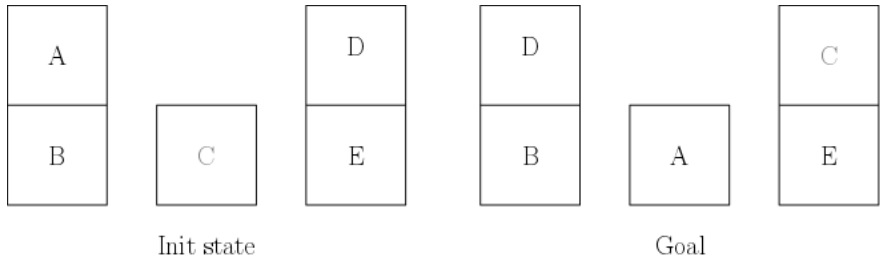


Figure 4: Instance for the Swapping Blocks World domain

Motivate all of yours decisions.
Write your answer to Question 3.2(b) here.

3.3 Solving Planning [6 marks]

1. Consider the encoding provided in the previous exercise. Provide:
   - an optimal plan solving the problem of Figure 4
   - an optimal relaxed plan according to the delete free relaxation Π^+
   - What are $h^+(I)$ and $h^*(I)$?
   - a partial ordered plan. For this you need to identify the pair $\langle \pi, \prec \rangle$ that is aimed at capturing the actions needed (first element of the pair), and the ordering relations ensuring only valid linearizations of the actions in π. You can ignore the *start* and the *end* dummy actions.

Write your answer to Question 3.3(b) here.

4 Reinforcement Learning [25 Marks]

4.1 True or False (5 marks)

a) [1.25/25] A policy is a function from actions to states.

b) [1.25/25] The optimal policy in an environment is always unique.

c) [1.25/25] Reinforcement learning is also known as stochastic planning.

d) [1.25/25] The Markov property states that the future is independent of the past given the present.

Mark your answers to Question 4.1 on the exam paper.

4.2 Value functions (8 marks)

a) [1/25] Write down the (mathematical) definition of a state-value function $V^{\pi}(s)$ for a discounted MDP.

Write your answer to Question 4.2(b) here.

b) [1/25] Define the optimal state-value function $V^{*}(s)$ for an MDP.

Write your answer to Question 4.2(b) here.

c) [1/25] Define the optimal action-value function $Q^{*}(s, a)$ via the optimal state-value function.

Write your answer to Question 4.2(c) here.

d) [1/25] Define the optimal policy in terms of the optimal action-value function.

Write your answer to Question 4.2(d) here.

e) [3/25] Let π_1 be defined as the greedy policy with respect to some π i.e. $\pi_1(s) = \arg\max_a Q^{\pi}(s,a)$. Show that if $V^{\pi_1}(s) = V^{\pi}(s)$ for all s, i.e. the new policy does not improve, then π_0 and π_1 must be optimal policies. Hint: use the Bellman (optimality) equation.

Write your answer to Question 4.2(e) here.

4.3 Example Domain (5 marks)

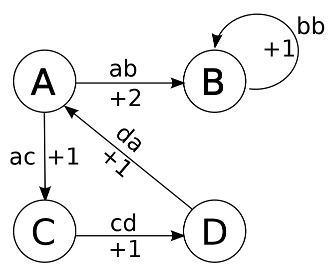


Figure 5: Question 4: MDP problem

Figure 5 defines a deterministic MDP with states $S = A, B, C, D$, actions $A = ab, bb, ac, cd, da$ and rewards $R = +1, +2.$. For the following questions assume that $\gamma = 1/2$.

a) [2/25] Calculate the value of state A under the optimal policy.

Write your answer to Question 4.3(a) here.

b) [3/25] Assume instead that a uniform random policy was used (at state A, choose ac or ab with probability 1/2). What is the value of state A under this policy?

Write your answer to Question 4.3(b) here.

4.4 Temporal Difference Learning (7 marks)

a) [2/25] What is the primary difference between TD(0) and SARSA?

Write your answer to Question 4.4(a) here.

b) [2/25] What is the primary difference between SARSA and Q-learning? Illustrate the difference by providing an example where their performances would differ.

Write your answer to Question 4.4(b) here.

c) [3/25] In the example in Figure 5, assume that you have seen the trajectory *A,C,D,A,B*. If you have a learning rate of 0.5, and initialise $V(s) = 10\,\forall s$, what will $V(A)$ be after performing a TD(0) update for the trajectory.

Write your answer to Question 4.4(b) here.

Additional answers to question: ___.

Additional answers to question: ___.

Additional answers to question: ___.

Additional answers to question: ___.

Additional answers to question: ___.

Additional answers to question: ___.

Additional answers to question: ___.