THE UNIVERSITY OF HONG KONG

FACULTY OF ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE

COMP7404 Computational Intelligence and Machine Learning

Date: May 23, 2016 **Time: 6:30pm - 8:30pm**

Only approved calculators as announced by the Examinations Secretary can be used in this examination. It is the candidates' responsibility to ensure that their calculator operates satisfactorily, and candidates must record the name and type of the calculator used on the front page of the examination script.

Calculator Brand: ____________________ *Calculator Model:* ____________________

Write your University No. on every page.

Answer all six questions in the space provided.

Special Note: Candidates are permitted to bring to the examination ONE sheet of A4-sized paper with printed or written notes on both sides.

Question No.	Weight (in %)	Your Mark (in %)
1	18	
2	18	
3	14	
4	20	
5	22	
6	8	
Total	100	

1. True / False Questions

For each True / False question, circle the correct answer and optionally also provide an explanation. A correct answer is worth 1 mark, a missing answer is worth 0 marks and a wrong answer with no explanation is worth -1 marks.

(a) [*True* or *False*] The heuristic $h(n) = 0$ is admissible for every search problem.

(b) [*True* or *False*] The heuristic $h(n) = 1$ is admissible for every search problem.

(c) [*True* or *False*] The heuristic $h(n) = c(n)$, where $c(n)$ is the true cheapest cost to get from the node n to a goal state, is admissible for every search problem.

(d) [*True* or *False*] Uniform-cost search will never expand more nodes than A*-search.

(e) [*True* or *False*] Depth-first search will always expand more nodes than breadth-first search for the same search problem.

(f) [*True* or *False*] Alpha-beta pruning prunes the same number of subtrees independent of the order in which successor states are expanded.

(g) [*True* or *False*] Minimax search with alpha-beta pruning generally requires more run-time than minimax without pruning on the same game tree.

(h) [*True* or *False*] The starting position for chess has a value of either 1 (white wins), -1 (black wins), or 0 (draw).

(i) [*True* or *False*] Assume *Agent* 1 has a utility function U_1 and *Agent* 2 has a utility function U_2. If $U_1 = k_1 U_2 + k_2$ with $k_1 > 0, k_2 > 0$ then *Agent* 1 and *Agent* 2 have the same preferences.

(j) [*True* or *False*] Assume $A \succ B$, $D \succ A$, $B \succ L$, where $L = [0.5, C; 0.5, D]$. With rational preferences it follows that $B \succ C$.

(k) [*True* or *False*] If the policy as converged in value iteration, the values must have converged as well.

(l) [*True* or *False*] If the values have converged in value iteration, the policy must have converged as well.

(m) [*True* or *False*] For an MDP if we just change the reward function R the optimal policy is guaranteed to remain the same.

(n) [*True* or *False*] If the only difference between two MDPs is the value of the discount factor then they must have the same optimal policy.

(o) [*True* or *False*] For an infinite horizon MDP with finite number of states and actions with a discount factor γ, with $0 < \gamma < 1$, policy iteration is guaranteed to converge.

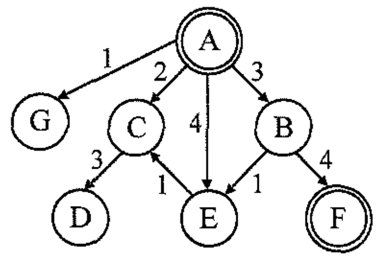
(p) [*True* or *False*] Q-Learning can learn the optimal Q-function Q^* without ever executing the optimal policy.

(q) [*True* or *False*] If an MDP as a transition model T that assigns non-zero probability for all tuples T(s,a,s') then Q-Learning will fail.

(r) [*True* or *False*] Q-learning is only guaranteed to converge to the optimal q-values if the underlying MDP is deterministic.

2. Search

(a) Consider the following search graph. A is the start state and F is the goal state. Assume that children are added to the frontier in alphabetical order.

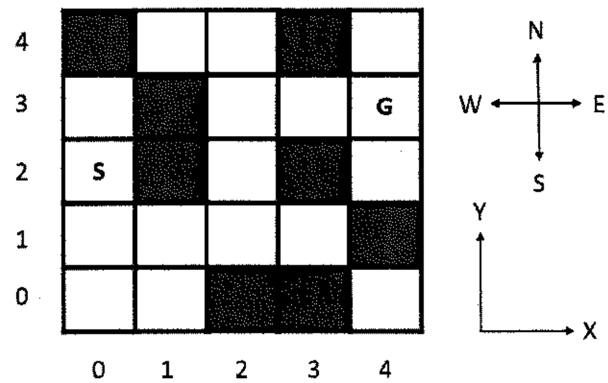


i. Write down the order in which states will be explored by depth-first tree search. *(4 marks)*

ii. Write down the order in which states will be explored by breadth-first tree search. *(4 marks)*

iii. Write down the order in which states will be explored by uniform-cost graph search. If there is a tie use *last-in first-out.* *(4 marks)*

(b) Consider the following maze in which the *Start* block is located at $(0, 2)$, and the *Goal* block at $(4, 3)$. The agent can move vertical and horizontal. Assume that children are added to the frontier in the following order: North, East, South, West. Dark blocks represent a wall through which the agent cannot move.



Solve the maze using the A* graph search algorithm. Use the Manhattan distance as the heuristic function. Write down the explored set in the state-explored order. If there is a tie use *last-in first-out.* *(6 marks)*

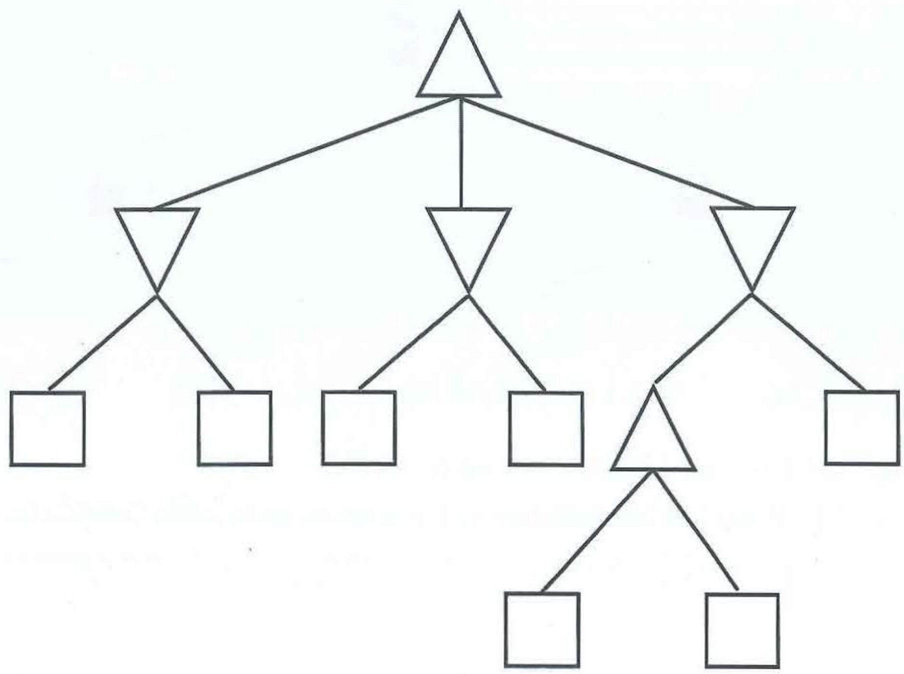
3. Adversarial Search

For each of the following minimax game trees, fill in the leaf nodes with 0 or 1 as utility values. For the first tree fill in the leaves so that as many leaves will be pruned by alpha-beta as possible. For the second tree fill in the leaves so that no leaves are pruned. *(14 marks)*

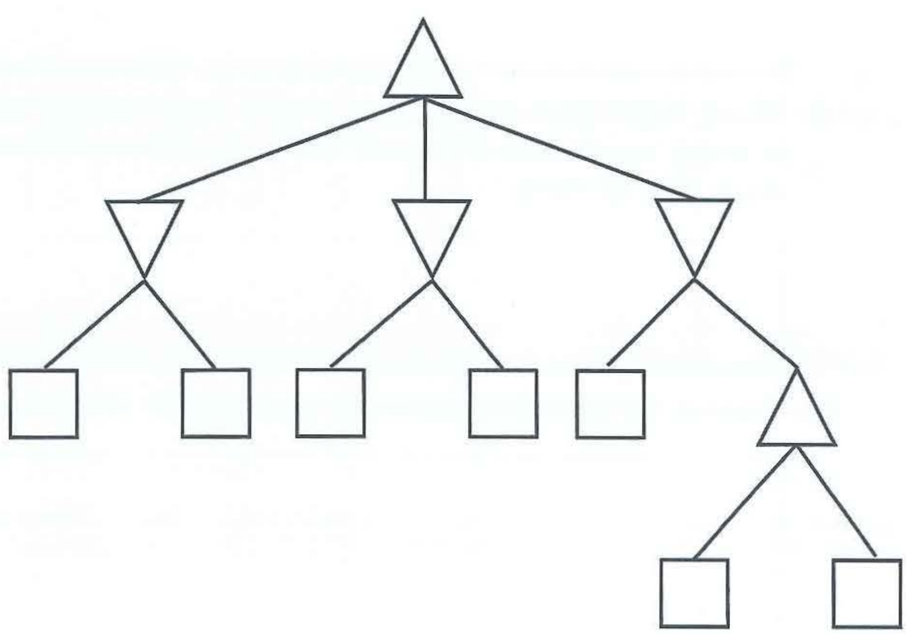
Assume:

(a) left to right traversal while pruning, and

(b) leaf values are drawn from $[-\infty, \infty]$.



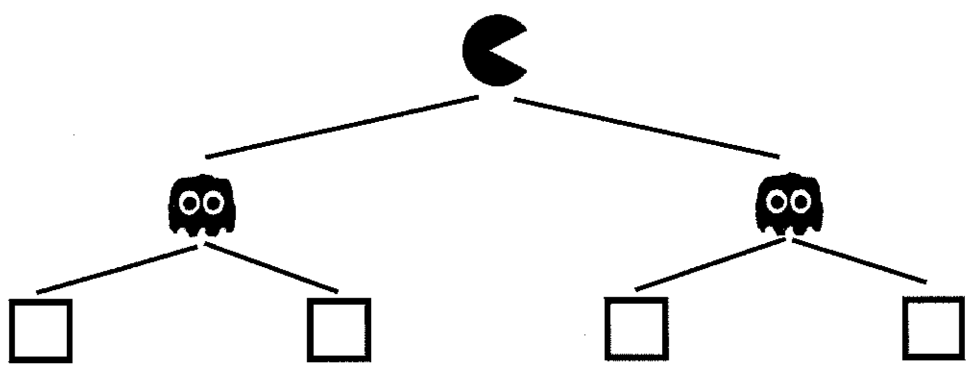
(i)



(ii)

4. Utilities

Pacman is buying a raffle ticket (a raffle ticket is a type of gambling ticket) from the ghost ticket vendor. There are two ticket types: A and B, but there are multiple specific tickets of each type. Pacman buys a ticket type, but the ghost will then choose which specific ticket Pacman will receive. Pacman's utility for a given raffle ticket is equal to the utility of the lottery of outcomes for that raffle ticket. For example, ticket $R_{A,2}$ corresponds to a lottery with equal chances of yielding 10 and 0, and so $U(R_{A,2}) = U([0.5, 10; 0.5, 0])$. Pacman, being a rational agent, wants to maximize his expected utility, but the ghost may have other goals! The outcomes are illustrated below.



$R_{A,1} : [1, \$6]$ $\quad R_{A,2} = [0.5, \$10; 0.5, \$0]$ $\quad R_{B,1} : [1, \$7]$ $\quad R_{B,2} : [\frac{1}{3}, \$3; \frac{1}{3}, \$6; \frac{1}{3}, \$9]$

(a) Let Pacman's utility for money be $U(m) = m$.

i. What are the utilities to Pacman of each raffle ticket? *(4 marks)*

ii. Which raffle ticket will Pacman receive under optimal play if the ghost is trying to minimize Pacman's utility (and Pacman knows the ghost is doing so)? *(2 marks)*

iii. What is the equivalent monetary value of raffle ticket $R_{B,2}$? *(2 marks)*

(b) Now, let Pacman's utility for money be $U(m) = m^2$.

i. What are the utilities to Pacman of each raffle ticket? *(4 marks)*

ii. The ghost is still trying to minimize Pacman's utility, but the Ghost mistakenly thinks that Pacman's utility is given by $U(m) = m$, and Pacman is aware of this flaw in the ghost's model. Which raffle ticket will Pacman receive? *(2 marks)*

iii. What is the equivalent monetary value of raffle ticket $R_{B,2}$? *(2 marks)*

(c) Pacman has a raffle ticket with distribution $[0.5, \$100; 0.5, \$0]$. A ghost insurance dealer offers Pacman an insurance policy where Pacman will get \$100 regardless of what the outcome of the ticket is. If Pacman's utility for money is $U(m) = \sqrt{m}$, what is the maximum amount of money Pacman would pay for this insurance? *(4 marks)*

5. Markov Decision Processes

Consider an MDP with transitional model and reward function as follows.

s	a	s'	T(s,a,s')	R(s,a,s')
A	1	A	0	8
A	1	B	1	2
A	2	A	1	1
A	2	B	0	2
A	3	A	0.5	4
A	3	B	0.5	2
B	1	A	0.5	10
B	1	B	0.5	8
B	2	A	1	0
B	2	B	0	4
B	3	A	0.5	2
B	3	B	0.5	4

Assume a discount factor of $\gamma = 0.5$.

(a) Calculate the following values of the MDP *(18 marks)*:

$V_0^*(A) =$ ____________________

$V_0^*(B) =$ ____________________

$Q_1^*(A, 1) =$ ____________________

$Q_1^*(A, 2) =$ ____________________

$Q_1^*(A, 3) =$ ____________________

$Q_1^*(B, 1) =$ ____________________

$Q_1^*(B, 2) =$ ____________________

$Q_1^*(B, 3) =$ ____________________

$V_1^*(A) =$ ________________________________

$V_1^*(B) =$ ________________________________

$Q_2^*(A, 1) =$ ________________________________

$Q_2^*(A, 2) =$ ________________________________

$Q_2^*(A, 3) =$ ________________________________

$Q_2^*(B, 1) =$ ________________________________

$Q_2^*(B, 2) =$ ________________________________

$Q_2^*(B, 3) =$ ________________________________

$V_2^*(A) =$ ________________________________

$V_2^*(B) =$ ________________________________

(b) Let $\pi_i^*(s)$ be the optimal action in state s with i time steps to go. Fill in the following table with values for $\pi_i^*(s)$. *(4 marks)*

s	$\pi_1^*(s)$	$\pi_2^*(s)$
A		
B		

6. Reinforcement Learning

Consider an MDP with three states, A, B and C; and two actions CW and CCW. We do not know the transition function or the reward function for the MDP, but instead, we are given samples of what an agent actually experiences when it interacts with the environment (although, we do know that we do not remain in the same state after taking an action). In this problem, instead of first estimating the transition and reward functions, we will directly estimate the Q function using Q-learning. Assume, the discount factor, $\gamma = 0.25$ and the step size for Q-learning, $\alpha = 0.75$. Let the current Q function, $Q(s, a)$, be

	A	B	C
CW	0.3	-0.5	-2.5
CCW	0.0	-1.3	-1.9

The agent encounters the following samples

s	a	s'	r
A	CCW	C	9.0
B	CW	A	1.0

Process the samples given above and write down all Q-values of the three states after both samples have been accounted for. *(8 marks)*

END OF PAPER

Please be reminded to write your university number on every page.

This page is intentionally left blank.
You may use it to draft your answers.