*University No.:*_____

THE UNIVERSITY OF HONG KONG

FACULTY OF ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE

Quiz 1

COMP3270 Artificial Intelligence

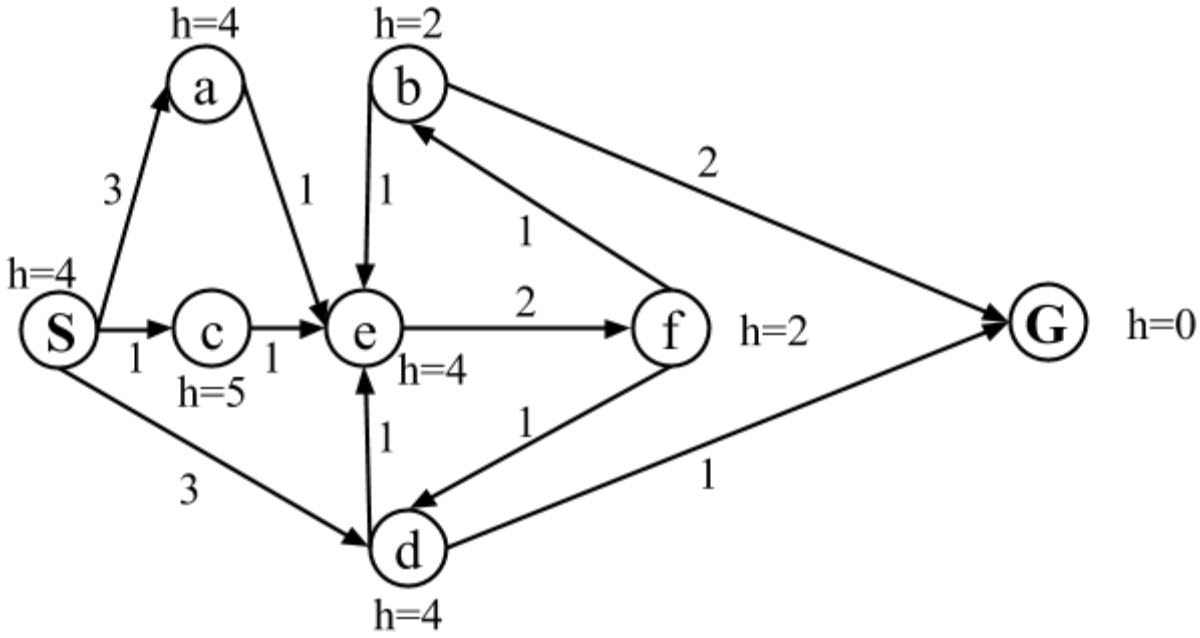
Only approved calculators as announced by the Examinations Secretary can be used in this examination. It is the candidates' responsibility to ensure that their calculator operates satisfactorily, and candidates must record the name and type of the calculator used on the front page of the examination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

Calculator Brand: _____ *Calculator Model:*_____

Answer questions in the space provided.
Write your University No. on every page.

Chapter	Max. Mark	Your Mark (examiner use only)
1	6	
2	4	
3	2	
4	4	
5	4	
Total	20	

1: Consider the following state space graph.



Let S be the start state and G be the goal state. Assume children are added into the frontier in alphabetical order. Write down the order of explored nodes for the following search algorithms / strategies.

1.1 (2 marks): Breadth-first search (BFS) - Tree Search Algorithm (TSA)

Sacdeeeg(G)

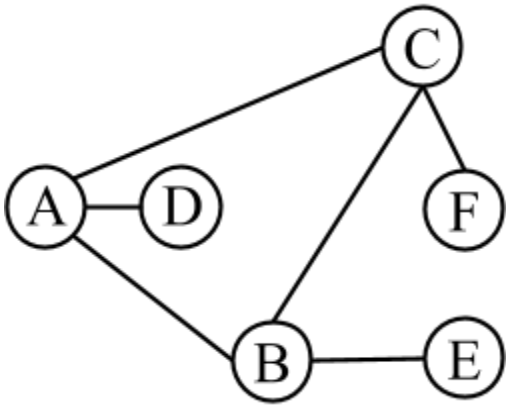
1.2 (2 marks): Uniform-cost search (UCS) - Graph Search Algorithm (GSA)
Use last in - first out as tiebreaker

Sceda (G)

1.3 (2 mark): A* Search - Graph Search Algorithm (GSA)
Use last in - first out as tiebreaker

Scefb (G)

2: The graph below is a constraint graph for a constraint satisfaction problem (CSP) that has only binary constraints. Initially, no variable has been assign.



For each of the following scenarios, write down all variables for which the specified filtering might result in their domain being changed.

2.1 (1 mark): A value is assigned to A. Which domains might change as a result of forward checking?

B C D

2.2 (1 mark): A value is assigned to A and then forward checking is run for A. Then a value is assigned to B. Which domains might change as a result of running forward checking for B?

C E

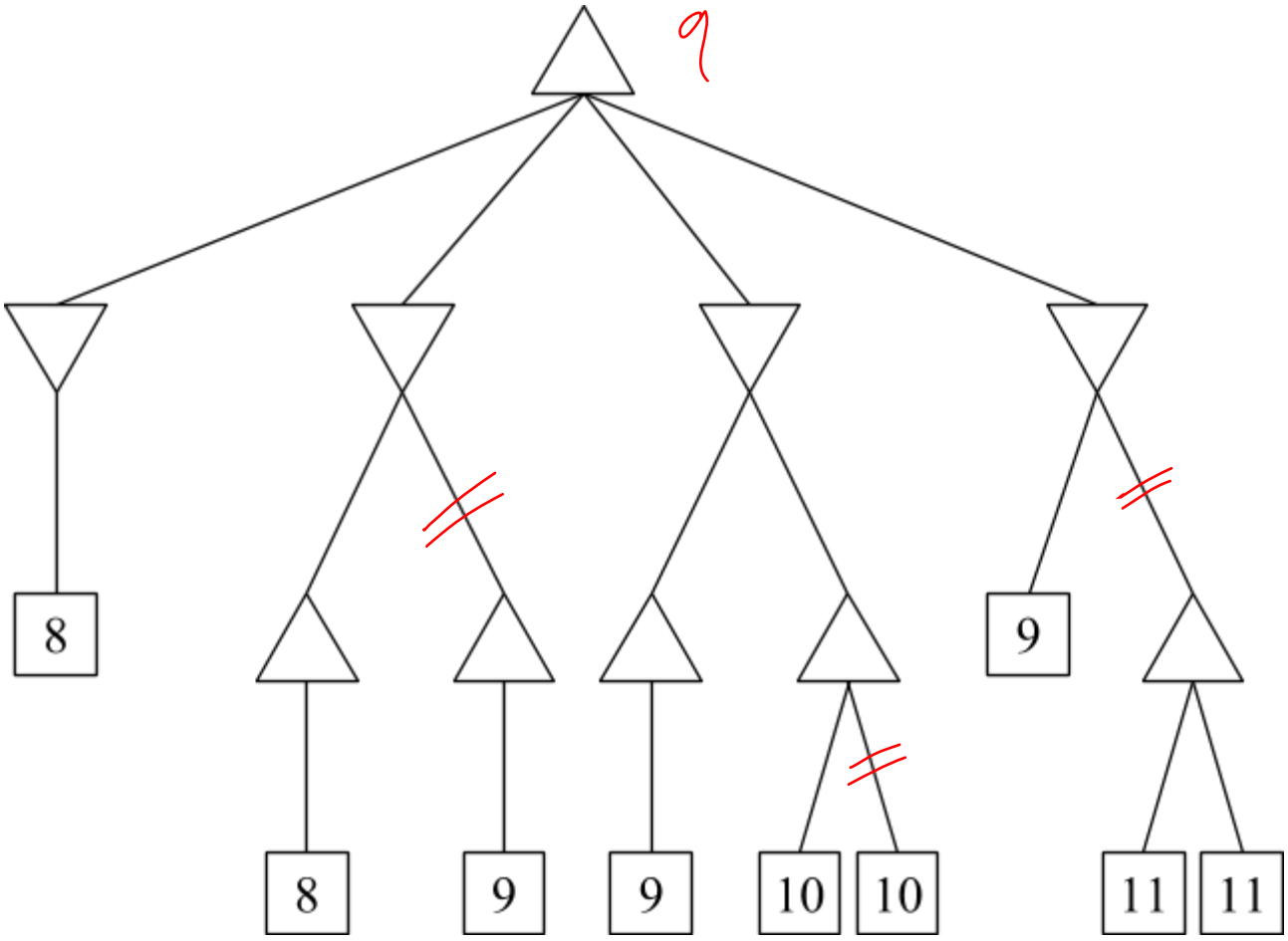
2.3 (1 mark): A value is assigned to A. Which domains might change as a result of enforcing arc consistency?

B C D E F

2.4 (1 mark): A value is assigned to A and then forward checking is run for A. Then a value is assigned to B. Which domains might change as a result of enforcing arc consistency after the assignment to B?

C E F

3 (2 mark): For the game tree shown below, indicate which branches will be pruned by alpha-beta pruning and write down the value at the root of the tree.



4 (4 marks): Consider the following Grid World introduced in our lecture where an agent may move north, south, east or west and exit states are a4 and b4.

-1.11	-1.10	-0.94 *	+1	a
-1.11		-1.10	-1	b
-1.11	-1.11	-1.11	-1.10	c
1	2	3	4	

Actions are unreliable: 80% of time each action achieves the desired effect and 20% of the time the action moves the agent at right angles (with equal probability) to the intended directions. If the agent runs into a wall, it stays in the same square. After running value iteration we have obtained the values shown in the figure. Note that values have been rounded to two decimal places. Let the discount factor $\gamma = 0.1$ and the reward $R(s) = -1$.

Calculate the Q-values of the state a3 and perform policy extraction for the state.

$Q(a_3, North) = 0.8 * (-1 - 0.94) + 0.1 * (-1 + 0.1) + 0.1 * (-1 - 0.11) = -1.0762$

$Q(a_3, South) = 0.8 * (-1 - 0.11) + 0.1 * (-1 + 0.1) + 0.1 * (-1 - 0.11) = -1.0762$

$Q(a_3, East) = 0.8 * (-1 + 0.1) + 0.1 * (-1 - 0.094) + 0.1 * (-1 - 0.11) = -0.9494$

$Q(a_3, West) = 0.8 * (-1 - 0.11) + 0.1 * (-1 - 0.094) + 0.1 * (-1 - 0.11) = -1.1084$

$\pi(a_3) = East$

5 (4 marks): 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 $\gamma = 0.75$ and the learning rate for Q-learning $\alpha = 0.5$. Let the current Q function, Q(s, a), be:

	A	B	C
CW	0.3	-0.5	-2.5
CCW	0.1	-1.3	-1.9

2.00015625

3.3375

The agent encounters the following samples:

s	a	s'	r
A	CCW	C	8.0
B	CW	A	2.0

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

$Q(A, CCW) = 0.5 * 0.1 + 0.5 * [8 + 0.75 * (-1.9)]$

$= 3.3375$

$Q(B, CW) = 0.5 * -0.5 + 0.5 * [2 + 0.75 * 3.3375]$

$= 2.00015625$

END OF PAPER

You may use this page to draft your answer