UID: ____________________

THE UNIVERSITY OF HONG KONG

FACULTY OF ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE

Quiz 1

Time: 12:30am - 1:20pm
Date: 18 October 2022

COMP3270 Artificial Intelligence

- Write your University No. at the top of all pages
- This is a closed book examination
- Use of a calculator is allowed. Only approved calculators as announced by the Examinations Secretary can be used in this examination. It is 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
- Answer ALL questions in the space provided

		Your Mark
Total	40	

Question 1

1.1 [4 marks]: Draw a graph that represents a search problem. Write down costs for all arcs and heuristics for each state such that the search problem is admissible and not consistent. Use as few states as possible.

1.2 [2 marks]: Do the same again as in the previous question. This time make sure your heuristic is consistent and not admissible.

1.3: Consider a knight on a chessboard. The knight is a piece represented by a horse's head and neck. It may move two squares vertically and one square horizontally or two squares horizontally and one square vertically. In the following, assume we are operating on an unbounded chessboard.

1.3.1 [3 marks]: Consider the problem of moving a single knight from position x to y. Write down the branching factor for this search problem's state space. Explain.

1.3.2 [3 marks]: Consider the problem of moving k knights from position $x_1 \ldots x_k$ to $y_1 \ldots y_k$ in the fewest number of moves. Knights cannot occupy the same square. Write down the maximum branching factor assuming only a single knight may move at a time. Explain.

1.3.3 [3 marks]: Suppose h_i is an admissible heuristic for the problem of moving knight i from x_i to y_i just by itself. Which of the following heuristics are admissible for the k-knight problem? Explain.

a) $\min(h_1, \ldots, h_k)$ b) $\max(h_1, \ldots, h_k)$ c) $\text{sum}(h_1, \ldots, h_k)$

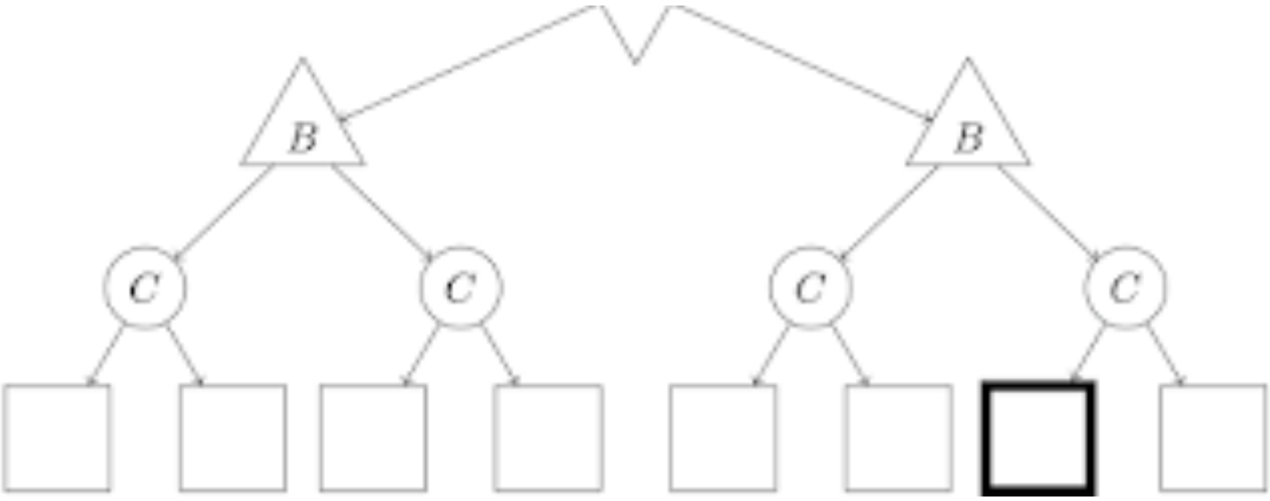
1.3.4 [3 marks]: With reference to 1.3.3, which one is the best heuristic? Explain.

1.4: In this question, player A is a minimizer, player B is a maximizer, and C represents a chance node. All children of a chance node are equally likely. Consider a game tree with players A, B, and C. In lecture, we considered how to prune a minimax game tree - in this question, you will consider how to prune an expinimax game tree (like a minimax game tree but with chance nodes). Assume that the children of a node are visited in left-to-right order.

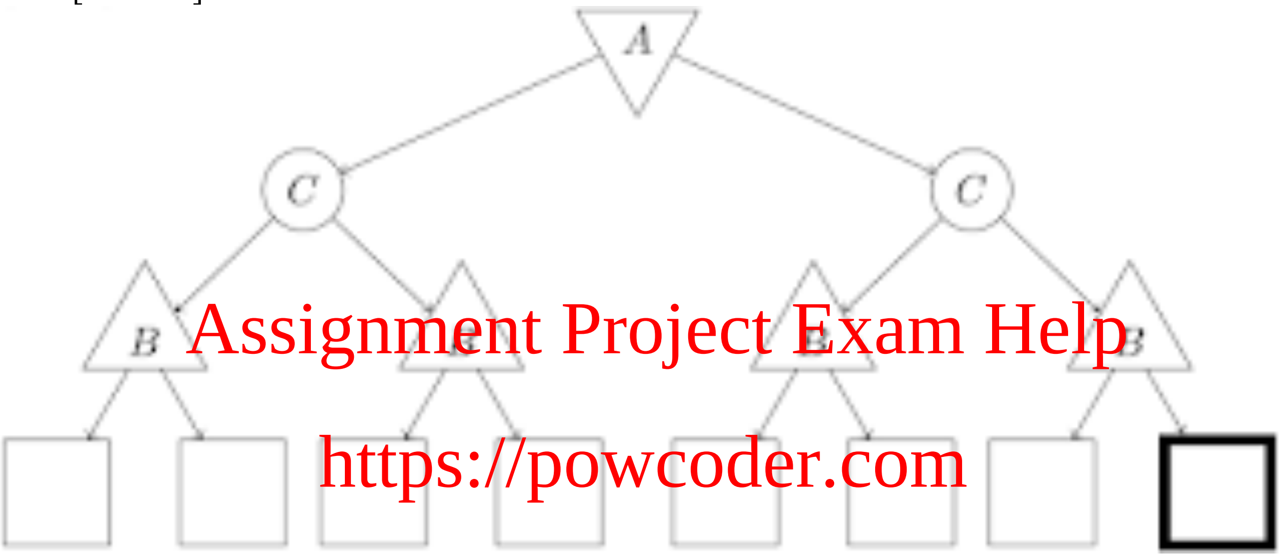
For each of the following game trees, give an assignment of terminal values to the leaf nodes such that the bolded node can be pruned, or write "not possible" if no such assignment exists. You may give an assignment where an ancestor of the bolded node is pruned (since then the bolded node will never be visited). Your terminal values must be finite and you should not prune on equality.

Important: The α-β pruning algorithm does not deal with chance nodes. Instead, for a node n, consider all the values seen so far, and determine whether you can know without looking at the node that the value of the node will not affect the value at the top of the tree. If that is the case, then n can be pruned.

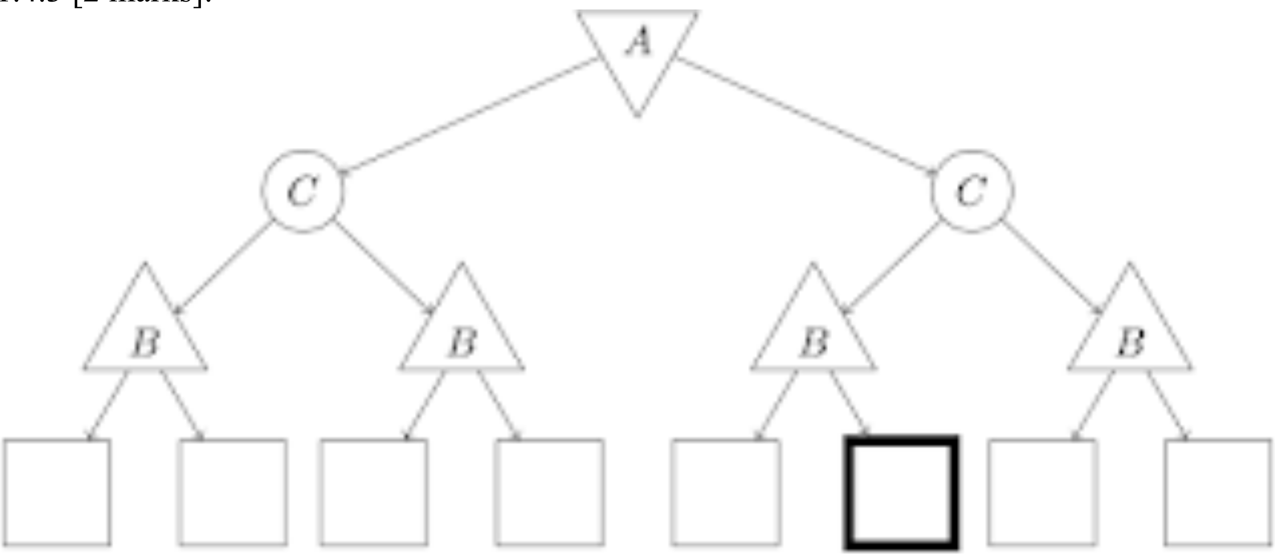
1.4.1 [2 marks]:



1.4.2 [2 marks]:



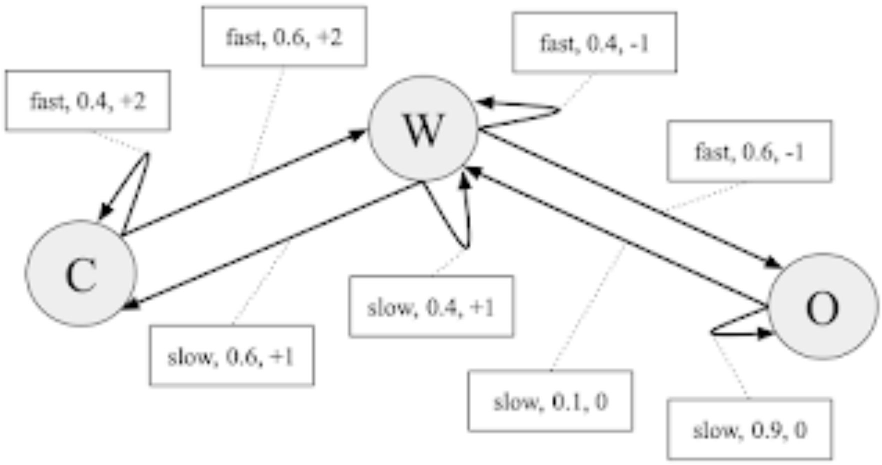
1.4.3 [2 marks]:



Question 2
Consider value iteration formula discussed in class

$$V_{k+1}(s) \leftarrow \max_a \sum_{s'} T(s,a,s')[R(s,a,s') + \gamma V_k(s')]$$

and the modified overheating car problem given as follows.



The rectangles denote *action*, *probability*, *reward*. Let the discount factor be 0.8, determine V_2 for states Cool (C) and Warm (W). Show your work.

Question 3

3.1 [6 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 (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	0.5	-1.5	-2.5
CCW	1.5	-1.5	-2

The agent encounters the following samples:

s	a	s'	r
A	CCW	C	2.5
B	CW	A	1.5

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

3.2 [2 marks]: In reinforcement learning, why can it be useful to sometimes act in a way which is believed to be suboptimal?

END OF PAPER