CSE 415 Winter 2020 Assignment 4

Last name: Mor First name: Saasha

Due Friday night February 7 via Gradescope at 11:59 PM. You may turn in either of the following types of PDFs: (1) Scans of these pages that include your answers (handwriting is OK, if it's clear), or (2) Documents you create with the answers, saved as PDFs. When you upload to GradeScope, you'll be prompted to identify where in your document your answer to each question lies.

Do the following five exercises. These are intended to take 20-25 minutes each if you know how to do them. Each is worth 20 points.

---

1 Blind Search

Sudoku is a popular type of puzzle that has many variations. One variation that is usually easier to solve than standard Sudoku is called Sudoku X 6 . The goal is to fill in the blank squares to get an array in which each of the rows, columns, two main diagonals, and six "regions" has each of the numbers 1 through 6 in it. The regions are 2 by 3 blocks, shown with bold outlines in the pictures below. (Standard Sudoku has 3 by 3 blocks.)

	1			4	
	2			5	
	3			6	

3	4	2	5	1	6
6	1	5	2	4	3
5	6	4	1	3	2
1	2	3	6	5	4
4	5	6	3	2	1
2	3	1	4	6	5

(a) (5 points) Describe a possible state representation for this puzzle, using either English, pseudocode or Python.

A 2D array with 36 elements (6x6) which represents each square on the Sudoku board

(b) (10 points) Provide pseudocode for a method successor(s) that takes a state s and returns a list of its successors.

```
def successor(s):
    list = []
    for i in range(7):
        for sq in s state :
            if sq is empty add
            state with i in 2 to list
    return list
```

(c) (5 points) Provide pseudocode for a method is_goal(s) that takes a state s and returns True of s is a goal state.

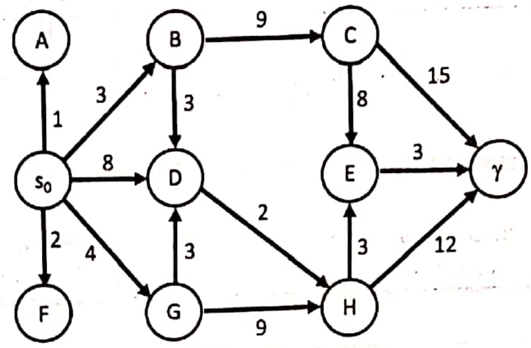
```
def is-goal(s) :
    return ( s = = goal state )
```

↓ which is the correct solution represented as a 2D array

2 Heuristic Search

(a) (5 points) Describe some of the challenges and trade-offs that need to be considered when selecting a heuristic.

The challenge is finding a heuristic that is admissible and consistent. We have to account for trade-offs that might come with choosing a value that is too high or low i.e. efficiency (low is not) or overestimating distance to goal. (high h_0)



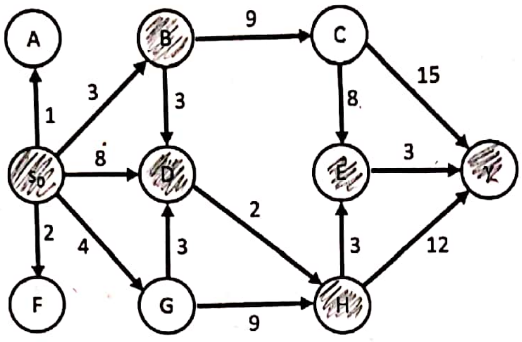
state (s)	s_0	A	B	C	D	E	F	G	H	γ
heuristic $h_1(s)$	14	15	4	10	3	2	16	10	5	0
heuristic $h_2(s)$	14	15	10	10	7	2	16	10	5	0
heuristic $h_3(s)$	14	15	12	10	7	2	16	10	5	0

(b) (5 points) Which heuristics (h_1, h_2, h_3) shown above are admissible? h_1, h_2

(c) (5 points) Which heuristics (h_1, h_2, h_3) shown above are consistent? —

(d) (5 points) Which of the 3 heuristics shown above would you select as the best heuristic to use with A* search, and why? Refer to consistency/admissibility in your justification.

Though the optimal heuristic is both consistent & admissable which none of the options possess.
I would choose h_2 since it is fairly efficient & doesn't overestimate distance.



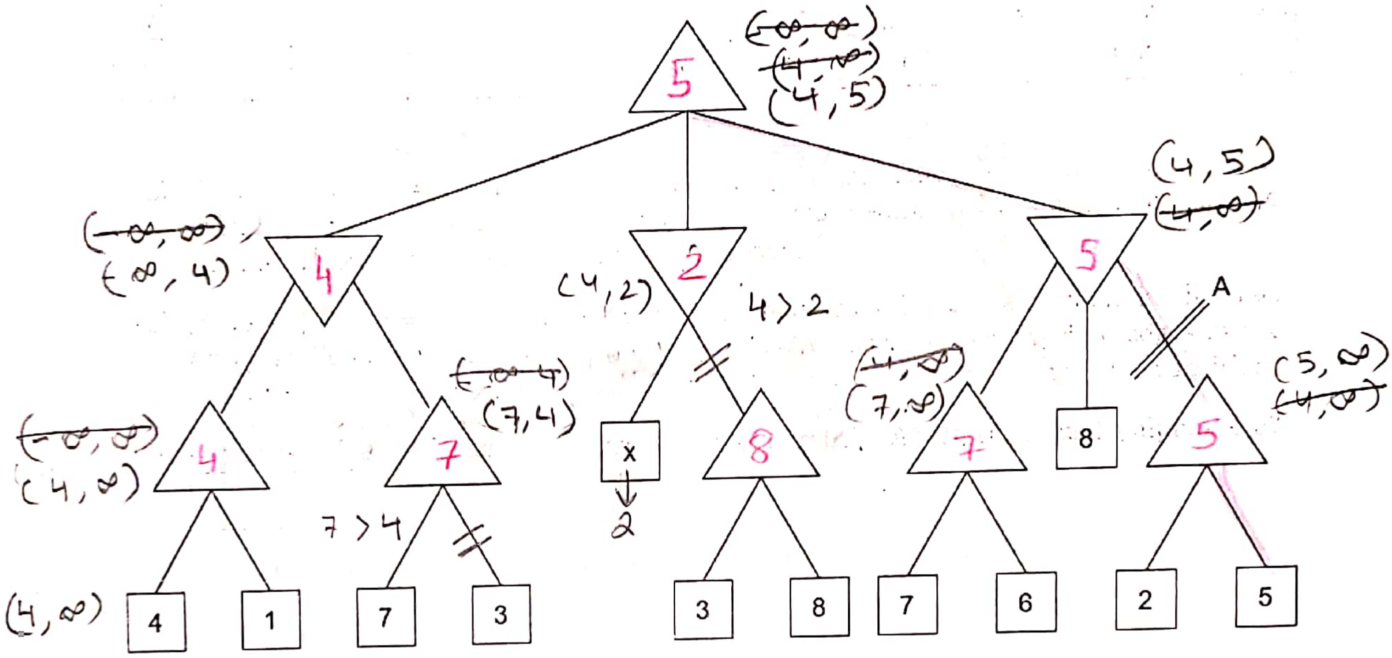
state (s)	s_0	A	B	C	D	E	F	G	H	γ
heuristic $h_3(s)$	6	8	5	5	3	1	8	5	2	0

(e) (5 points) Referring back to the graph again, trace out the path that would be followed in an A* search, given the heuristics provided above. As you trace the path, complete the table below, indicating which nodes are on the open and closed lists, along with their 'f' values:

	Open	Closed
Starting A^* search	[s_0, 6]	empty
s_0	[(B,8),(A,9),(G,9), (F,10), (D,11)]	[(S0,6)]
B	[(A,9),(G,9),(D,9), (F,10), (C,17)]	[(S0,6),(B,8)]
A	[(G,9),(D,9) (F,10)(C,17)]	[(S0,6),(B,8),(A,9)]
G	[(D,9),(F,10),(H,15),(C,17)]	[(S0,6) (B,8) (A,9), (G,9)]
D	[(F,10),(H,12),(C,17)]	[(S0,6) (B,8),(A,9) (G,9), (D,9)]
F	[(H,12) (C,17)]	[(S0,6) (B,8) (A,9),(G,9), (D,9), (F,10)]
H	[(E,14) (C,17) (γ,22)]	[(S0,6),(B,8) (A,9),(G,9), (D,9), (F,10),(H,12)]
E	[(γ,16),(C,17)]	[(S0,6) (B,8) (A,9) (G,9), (D,9) (F,10) (H,12) (E,14)]
γ	[(C,17)]	[(S0,6),(B,8) (A,9),(G,9), (D,9),(F,10) (H,12) (E,14)]

3 Adversarial Search

In the following minimax game tree, Δ represent maximizing nodes and ∇ represent minimizing nodes while □ nodes represent static evaluations of the states. Note that these values are not known to the search algorithm until it visits and explicitly evaluates the state. For the purposes of this question, the nodes in the tree should be processed from left to right. You may ignore the double slash (//) pruned branch labeled "A" until part c.



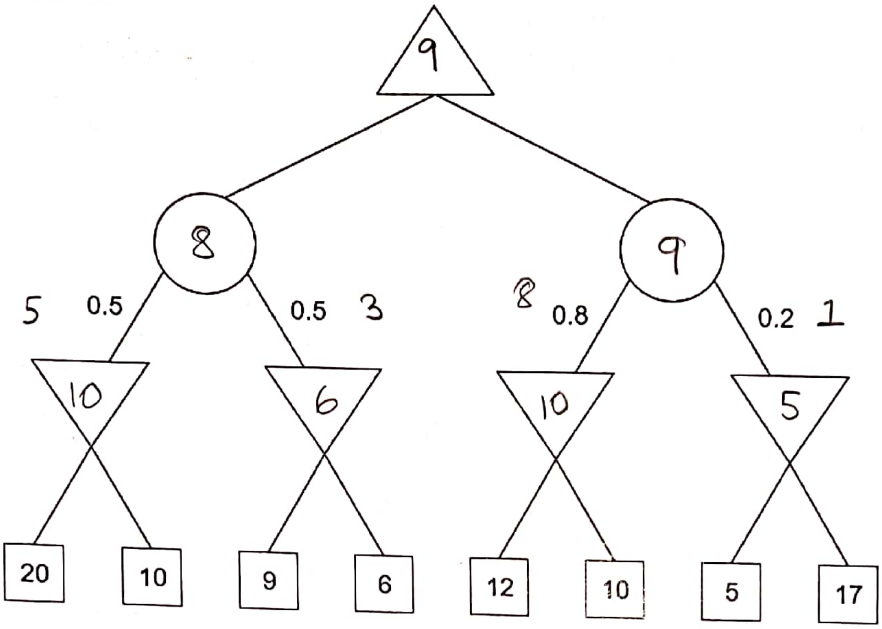
(a) (5 points) Let $x = 2$ in the static evaluation node containing the variable in the tree. Fill in the nodes in the tree with the correct values selected by the maximizing and minimizing players during the minimax algorithm. (middle values)

(b) (5 points) Let $x = 2$ in the static evaluation node containing the variable in the tree. Apply alpha-beta pruning to the tree. Specifically, mark the appropriate edges in the tree with a double slash (//) to indicate which nodes were pruned using the algorithm.

(c) (5 points) What is the smallest value of x where branch "A" is pruned using alpha-beta?

$\boxed{7}$

Consider the following expectimax game tree. Note that the new ○ nodes represent expectation nodes and the probability of their successors are denoted on the outgoing edges of these nodes.



(d) (5 points) Fill in the nodes in the tree with the correct values selected by the maximizing and minimizing players during the expectimax algorithm.

4 Markov Decision Processes

Consider the following game. In each turn you have a choice of rolling a special die, or stopping the game. The die is biased - every time you roll, it produces 1, 3, 5 or 6 with equal probability. No other values are possible (It's a tetrahedral die.) At any point of time, you can either roll or stop if the total "score" (obtained by adding the values on the die from every rolling) is less than 7. If the "score" reaches or exceeds 7, you "go bust" and go to the final state, accruing zero reward.

When in any state other than the final state, you are allowed to take the stop action. When you stop, you reach the final state and your reward is the total "score" if it is less than 7.

Note: there is no direct reward from rolling the dice (or we could say that there is a reward but it's always 0). The only non-zero reward comes from explicitly taking the stop action. Discounting or not should not matter in the MDP for this game, but for the record, we assume no discounting (i.e., $\gamma = 1$).

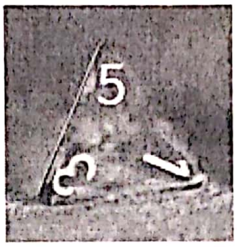


Figure 1: The value of a tetrahedral die like this, after a roll, is at the top, here 5, which should show equally well on any of the three faces that touch the top vertex.

(a) (6 points) Write down the states (in any order) and actions for this MDP. (Hint: there are 8 states in total and each should correspond to a numeric value except the initial and final states)

Actions → a_1 = Roll 1, a_2 = 3, a_3 = 5 or a_4 = 6 · or a_0 = stop

states ⇒ S_0 (initial), $\underbrace{S_1, S_2, S_3, S_4, S_5, S_6}_{\text{sums}}$, S_7 (final state)

(b) (10 points) Give the full transition function $T(s, a, s')$. Here s is a current state, a is an action, and s' is a possible next state when a is performed in s. Assuming your states are s_0, s_1, s_2, s_3 etc., and actions are a_0, a_1 etc., some examples of how you should write the function are as follows:

$$T(s, a, s') = \langle\text{value}\rangle;\ s = s_0,\ s' \in \{s_1, s_2, s_3, \ldots\}$$

$$T(s_0, a_1, s_1) = \langle\text{value}\rangle$$

$T(s, a, s') = 0.25$ → probability of rolling a number on a tetrahedral dice $(s = s_0)$

$a \rightarrow \{a_1, a_2 \cdots\}$ not a_0

$s \rightarrow \{s_1, s_2 \cdots\}$ not s_7

Example: $T(s_0, a_1, s_1) = 0.25$

(c) (2 points) Give the full reward function $R(s, a, s')$.

(1) $R(s_0, a_2, s_3) = 3.$

(2) $R(s_0, a_3, s_5) = 5$

(3) $R(s_1, a_1, s_2) = 2$

(4) $R(s_1, a_2, s_4) = 4$

(5) $R(s_0, a_4, s_5) = 6$

(6) $R(s_1, a_3, s_6) = 6$

(7) $R(s, a, s') = -10000$ (forfeit) where $s' = s_7$ and $s = s_0$ and $a = a_0$ or $a \neq a_0$

(8) $R(s, a, s') = 1$ $(s' = s_1)$

(9) $R(s, a, s') = 2$ where $s' = s_2$

(10) $R(s, a, s') = 3$ $(s' = s_3)$

(11) $R(s, a, s') = 4$ $(s' = s_4)$

(12) $R(s, a, s') = 5$, $s' = s_5$

(13) $R(s, a, s') = 10000$ $(s' = s_6)$ Best score

(d) (2 points) What is the optimal policy? There is no need to perform value iteration or use any fancy math; just write your answer in words.

The one where the agent avoids stopping the game to early & not overshooting 7. It should have preference to stop when reaching a sum total of 7.

5 Computing MDP State Values and Q-Values

Consider an MDP with two states s_1 and s_2 and transition function $T(s, a, s')$ and reward function $R(s, a, s')$. Let's also assume that we have an agent whose discount factor is $\gamma = 1$. From each state, the agent can take three possible actions, i.e., $a \in \{x, y, z\}$. The transition probabilities for taking each action and the rewards for transitions are shown below.

s	a	s'	$T(s,a,s')$	$R(s,a,s')$
s_1	x	s_1	1	0
s_1	x	s_2	0	0
s_1	y	s_1	0.5	4
s_1	y	s_2	0.5	1
s_1	z	s_1	0.4	0
s_1	z	s_2	0.6	10
s_2	x	s_1	0	0
s_2	x	s_2	1	0
s_2	y	s_1	0.9	2
s_2	y	s_2	0.1	8
s_2	z	s_1	1	1
s_2	z	s_2	0	0

Compute V_0 , V_1 and V_2 for states s_1 and s_2.

(a). $V_0(s_1) =$ 0
(b). $V_0(s_2) =$ 0
(c). $V_1(s_1) =$ 8.5
(d). $V_1(s_2) =$ 4.6
(e). $V_2(s_1) =$ 34.4
(f). $V_2(s_2) =$ 17.4

Now, compute Q_2 for states s_1 and s_2.
(g). $Q_2(s_1, x) =$ 8.5
(h). $Q_2(s_1, y) =$ 11
(i). $Q_2(s_1, z) =$ 14.5
(j). $Q_2(s_2, x) =$ 4.6
(k). $Q_2(s_2, y) =$ 7.2
(l). $Q_2(s_2, z) =$ 5.6

Finally, compute V_2 and Q_2 for both states, but with $\gamma = 0.5$.

(m). $V_2(s_1) =$ 29.75

(n). $V_2(s_2) =$ 15.1

(o). $Q_2(s_1, x) =$ 8.5

(p). $Q_2(s_1, y) =$ 11

(q). $Q_2(s_1, z) =$ 14.5

(r). $Q_2(s_2, x) =$ 4.6

(s). $Q_2(s_2, y) =$ 7.2

(t). $Q_2(s_2, z) =$ 5.6