Homework 2

Instructions

Points: Please see the points for each problem.
Submission Instructions: Please submit a PDF in Canvas.

Points

Question	Part	Points Possible	Points Earned
1. Minimax and α-β pruning	(a)	4	_____
	(b)	4	_____
2. MDP and Value Iteration	(a)	4	_____
	(b)	4	_____
	(c)	4	_____
3. Policy Gradient and PPO	(a)	3	_____
	(b)	3	_____
Total		**26**	_____

1. Minimax and α-β pruning Consider the minimax tree shown below for parts (a) and (b). [8]

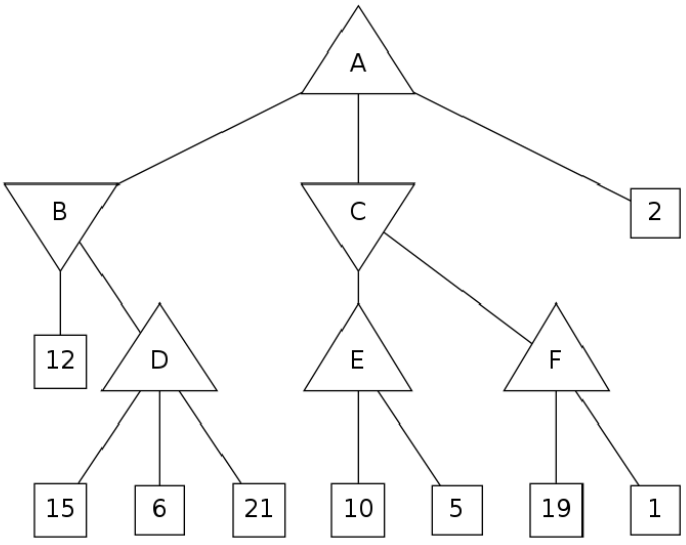

Figure 1: Minimax tree for Question 1.

(a) What value will the root node A have? [4]

(b) Cross off the nodes that are pruned by α-β pruning. Assume standard left-to-right traversal. If a non-terminal state (A, B, C, D, E, or F) is pruned, cross off the entire subtree. [4]

2. Pacman is using MDPs and Value Iteration to maximize his expected utility. He has the standard actions {North, East, South, West} unless blocked by an outer wall. There is a reward of 1 when eating a dot. The game ends when the dot is eaten. [12]

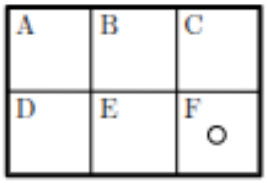

Figure 2: Grid for Question 2.

(a) Consider the grid where there is a single food pellet in the bottom-right corner (F) as shown in figure 2. The discount factor is $\gamma = 0.5$. There is no living reward. The states are the grid locations A, B, C, D, E, F. What is the optimal policy for each state? [4]

State s	Policy $\pi(s)$
A	
B	
C	
D	
E	

(b) What is the optimal value for the upper-left corner state A? [4]

(c) Using value iteration initialized with $V^0(\cdot) = 0$, at which iteration k does $V^k(A)$ first equal $V^*(A)$? [4]

3. Consider some of the newer RL algorithms such as Policy Gradient and Proximal Policy Optimization. [6]

(a) What issues with RL are they designed to solve? [3]

(b) How do they improve on previous RL algorithms? [3]