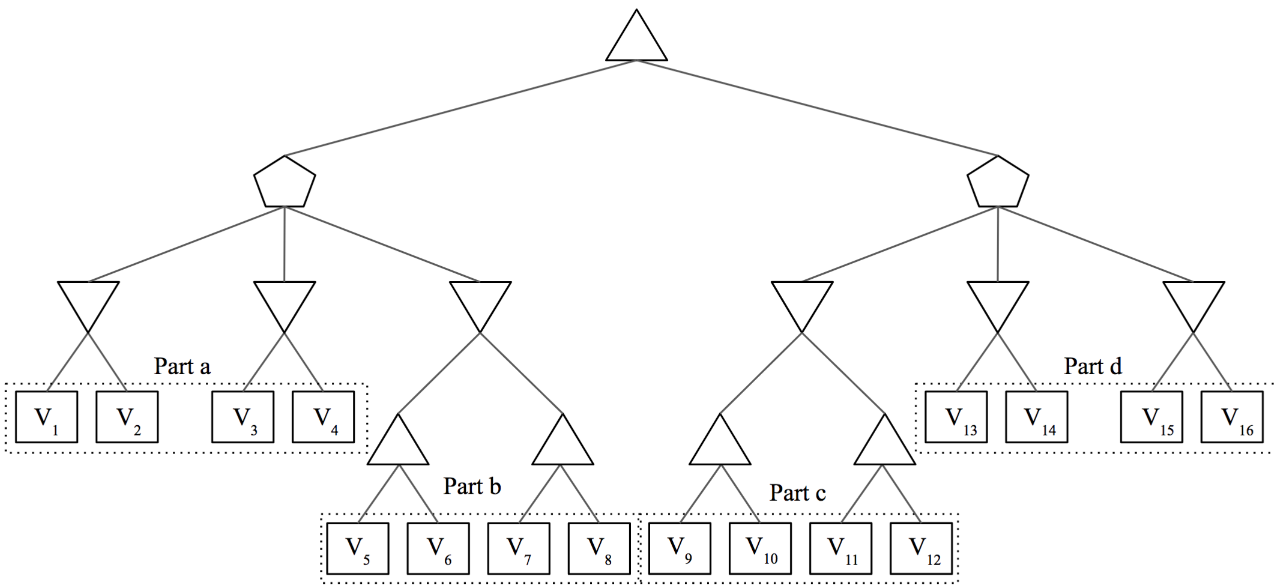
Q1. MedianMiniMax

You're living in utopia! Despite living in utopia, you still believe that you need to maximize your utility in life, other people want to minimize your utility, and the world is a 0 sum game. But because you live in utopia, a benevolent social planner occasionally steps in and chooses an option that is a compromise. Essentially, the social planner (represented as the pentagon) is a median node that chooses the successor with median utility. Your struggle with your fellow citizens can be modelled as follows:



There are some nodes that we are sometimes able to prune. In each part, mark all of the terminal nodes such that **there exists a possible situation** for which the node **can be pruned**. In other words, you must consider **all** possible pruning situations. Assume that evaluation order is **left to right** and all V_i's are **distinct.**

Note that as long as there exists ANY pruning situation (does not have to be the same situation for every node), you should mark the node as prunable. Also, alpha-beta pruning does not apply here, simply prune a sub-tree when you can reason that its value will not affect your final utility.

(a)
- ☐ V_1
- ☐ V_2
- ☐ V_3
- ☐ V_4
- ☐ None

(b)
- ☐ V_5
- ☐ V_6
- ☐ V_7
- ☐ V_8
- ☐ None

(c)
- ☐ V_9
- ☐ V_{10}
- ☐ V_{11}
- ☐ V_{12}
- ☐ None

(d)
- ☐ V_{13}
- ☐ V_{14}
- ☐ V_{15}
- ☐ V_{16}
- ☐ None

Q2. How do you Value It(eration)?

(a) Fill out the following True/False questions.

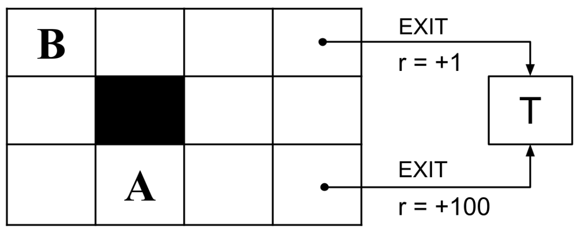
(i) ◯ True ◯ False: Let A be the set of all actions and S the set of states for some MDP. Assuming that $|A| \ll |S|$, one iteration of value iteration is generally faster than one iteration of policy iteration that solves a linear system during policy evaluation.

(ii) ◯ True ◯ False: For any MDP, changing the discount factor does not affect the optimal policy for the MDP.

The following problem will take place in various instances of a grid world MDP. Shaded cells represent walls. In all states, the agent has available actions $\uparrow, \downarrow, \leftarrow, \rightarrow$. Performing an action that would transition to an invalid state (outside the grid or into a wall) results in the agent remaining in its original state. In states with an arrow coming out, the agent has an *additional* action $EXIT$. In the event that the $EXIT$ action is taken, the agent receives the labeled reward and ends the game in the terminal state T. Unless otherwise stated, all other transitions receive no reward, and all transitions are deterministic.

For all parts of the problem, assume that value iteration begins with all states initialized to zero, i.e., $V_0(s) = 0\ \forall s$. **Let the discount factor be $\gamma = \frac{1}{2}$ for all following parts**.

(b) Suppose that we are performing value iteration on the grid world MDP below.



(i) Fill in the optimal values for A and B in the given boxes.

$V^*(A)$: ☐ $V^*(B)$: ☐

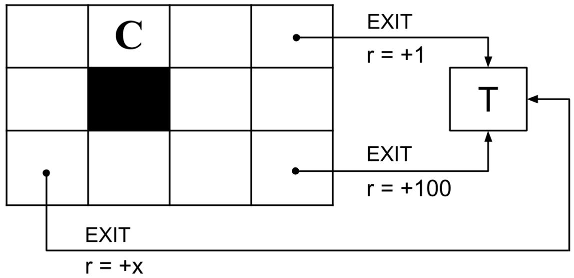
(ii) After how many iterations k will we have $V_k(s) = V^*(s)$ for all states s? If it never occurs, write "never". Write your answer in the given box.

☐

(iii) Suppose that we wanted to re-design the reward function. For which of the following new reward functions would the optimal policy **remain unchanged**? Let $R(s, a, s')$ be the original reward function.

☐ $R_1(s, a, s') = 10R(s, a, s')$
☐ $R_2(s, a, s') = 1 + R(s, a, s')$
☐ $R_3(s, a, s') = R(s, a, s')^2$
☐ $R_4(s, a, s') = -1$
☐ None

(c) For the following problem, we add a new state in which we can take the $EXIT$ action with a reward of $+x$.



(i) For what values of x is it *guaranteed* that our optimal policy π^* has $\pi^*(C) = \leftarrow$? Write ∞ and $-\infty$ if there is no upper or lower bound, respectively. Write the upper and lower bounds in each respective box.

$\boxed{} < x < \boxed{}$

(ii) For what values of x does value iteration take the **minimum** number of iterations k to converge to V^* for all states? Write ∞ and $-\infty$ if there is no upper or lower bound, respectively. Write the upper and lower bounds in each respective box.

$\boxed{} \leq x \leq \boxed{}$

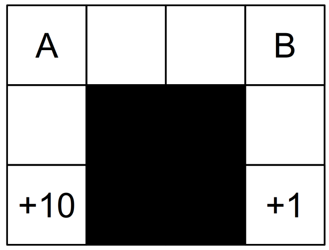
(iii) Fill the box with value k, the **minimum** number of iterations until V_k has converged to V^* for all states.

$\boxed{}$

Q3. MDPs: Value Iteration

An agent lives in gridworld G consisting of grid cells $s \in S$, and is not allowed to move into the cells colored black. In this gridworld, the agent can take actions to move to neighboring squares, when it is not on a numbered square. When the agent is on a numbered square, it is forced to exit to a terminal state (where it remains), collecting a reward equal to the number written on the square in the process.

Gridworld G



You decide to run value iteration for gridworld G. The value function at iteration k is $V_k(s)$. The initial value for all grid cells is 0 (that is, $V_0(s) = 0$ for all $s \in S$). When answering questions about iteration k for $V_k(s)$, either answer with a finite integer or ∞. For all questions, the discount factor is $\gamma = 1$.

(a) Consider running value iteration in gridworld G. Assume all legal movement actions **will always succeed** (and so the state transition function is deterministic).

(i) What is the smallest iteration k for which $V_k(A) > 0$? For this smallest iteration k, what is the value $V_k(A)$?

$k =$ ______________________ $V_k(A) =$ ______________________

(ii) What is the smallest iteration k for which $V_k(B) > 0$? For this smallest iteration k, what is the value $V_k(B)$?

$k =$ ______________________ $V_k(B) =$ ______________________

(iii) What is the smallest iteration k for which $V_k(A) = V^*(A)$? What is the value of $V^*(A)$?

$k =$ ______________________ $V^*(A) =$ ______________________

(iv) What is the smallest iteration k for which $V_k(B) = V^*(B)$? What is the value of $V^*(B)$?

$k =$ ______________________ $V^*(B) =$ ______________________

(b) Now assume all legal movement actions **succeed with probability** 0.8; with probability 0.2, the action fails and the agent remains in the same state.
Consider running value iteration in gridworld G. What is the smallest iteration k for which $V_k(A) = V^*(A)$? What is the value of $V^*(A)$?

$k =$ ______________________

$V^*(A) =$ ______________________