

**Project 3 - Question 9**

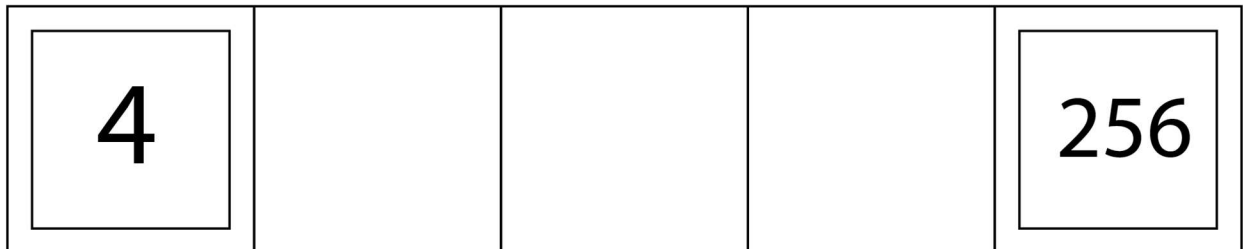
This non-programming problem is part of Project 3. Please add your answers to this document and submit your completed document along with your solution to the Pac-Man project.

The following problems take place in various scenarios of a 1D gridworld MDP.

In all cases double-rectangle states are exit states. From an exit state, the only action available is Exit, which results in the listed immediate reward and ends the game (by moving into a terminal state; not shown).

From non-exit states, the agent can choose either Left (L) or Right (R) actions, which move the agent in the corresponding direction. There are no living rewards; the only non-zero rewards come from exiting the grid. Throughout this problem, assume that value iteration begins with initial values  $V_0(s) = 0$  for all states.

Consider the following scenario:



**A**

**B**

**C**

**D**

**E**

Let the discount factor be  $\gamma = 0.75$ , and let transitions be deterministic. Fill in the missing values for each state following the value iteration algorithm in the following table:

Step $t$	$V_t(A)$	$V_t(B)$	$V_t(C)$	$V_t(D)$	$V_t(E)$
0	0	0	0	0	0
1					
2					
3					
4					

What are the actions for states B, C, and D for each time step, according to the best policy that corresponds to the values at step  $t$ ?

Time	$\pi_t(B)$	$\pi_t(C)$	$\pi_t(D)$
1			
2			
3			
4			