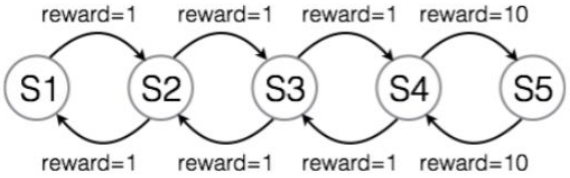
PROBLEM 4

Problem 5 Consider a reinforcement learning problem specified by the following Markov Decision Process (MDP).



We have five states representing steps along one direction. Call these states $S1$, $S2$, $S3$, $S4$, and $S5$. From each state, except the end states, we can move either left or right. The available actions in state $S1$ is just to move right while the action available in $S5$ is to move left. We can move left or right in each intermediate state. The reward for taking any action is 1 except when moving right from $S4$ or left from $S5$ which provide reward 10. Assume a discount factor $\gamma = 0.5$. Note that $\sum_{i=1}^{\infty} 0.5^i = 1$.

(a) **(5 points)** What is the optimal policy for this MDP? Specify action (L/R) to take in each state.

$$S1 : (\ \mathrm{R}\ ) \quad S2 : (\ \mathrm{R}\ ) \quad S3 : (\ \mathrm{R}\ ) \quad S4 : (\ \mathrm{R}\ ) \quad S5 : (\ \mathrm{L}\ ) \tag{12}$$

(b) Suppose we apply value iteration on this MDP. What is the value of state S3 after

(2 points) one value iteration

1

(2 points) two value iterations

$1 + 0.5\,10 = 6.0$

(3 points) ∞ number of value iterations

$1 + (\sum_{i=1}^{\infty} 0.5^i)10 = 11$

4) a) The optimal policy will be the one that moves towards higher rewards, so S5 will move left while all the other states will move right (towards S5).

b) one value iteration: 1
two value iterations: $1 + 0.5(10) = 6$
infinite value iterations: $1 + (\sum_{i=1}^{\infty} 0.5^i)(10) = 11$