For an MDP, the optimal policy is guaranteed to remain the same if we only change the reward function

1. True
2. False

For an MDP, the optimal policy is guaranteed to remain the same if we only change the reward function

1. True
2. False

Consider an MDP with 2 states A, B and 2 actions: “stay” stays at the current state and “move” moves to the other state. Let r be the reward function such that r(A) = 1, r(B) = 0. Let A be the start state and γ be the discounting factor.

Consider the “always move” policy π: π(A) = π(B) = move and an infinite sequence of A, B, A, B, … from this policy. What is the utility (i.e., the expected sum of discounted reward) of this sequence?

1. 0
2. 1 / (1 - γ)
3. 1 / (1 - γ^2)
4. 1

Consider an MDP with 2 states A, B and 2 actions: “stay” stays at the current state and “move” moves to the other state. Let r be the reward function such that r(A) = 1, r(B) = 0. Let A be the start state and γ be the discounting factor.

Consider the “always move” policy π: π(A) = π(B) = move and an infinite sequence of A, B, A, B, … from this policy. What is the utility (i.e., the expected sum of discounted reward) of this sequence?

1. 0
2. 1 / (1 - γ)
3. 1 / (1 - γ^2)
4. 1

Sequence: A, B, A, B, A, B, ...

Discounted rewards: 1, 0, γ^2, 0, γ^4, 0, γ^6, ...

Sum of discounted rewards: $1 + \gamma^2 + \gamma^4 + \gamma^6 + \cdots = 1/(1 - \gamma^2)$

In the above MDP, what is the optimal policy π^*? Assume A as the start state.

1. $\pi(A) = \pi(B) =$ move
2. $\pi(A) = \pi(B) =$ stay
3. $\pi(A) =$ stay, $\pi(B) =$ move
4. $\pi(A) =$ move, $\pi(B) =$ stay

In the above MDP, what is the optimal policy π^*? Assume A as the start state.

1. $\pi(A) = \pi(B) =$ move
2. $\pi(A) = \pi(B) =$ stay
3. $\pi(A) =$ stay, $\pi(B) =$ move
4. $\pi(A) =$ move, $\pi(B) =$ stay

Value iteration is guaranteed to converge if the discount factor (γ) satisfies $0 < \gamma < 1$.

1. True
2. False

Value iteration is guaranteed to converge if the discount factor (γ) satisfies $0 < \gamma < 1$.

1. True
2. False