Break & Quiz

Q 1.1 Which of the following statement about MDP is not true?

• A. The reward function must output a scalar value
• B. The policy maps states to actions
• C. The probability of next state can depend on current and previous states
• D. The solution of MDP is to find a policy that maximizes the cumulative rewards
Q 1.1 Which of the following statement about MDP is not true?

• A. The reward function must output a scalar value
• B. The policy maps states to actions
• C. The probability of next state can depend on current and previous states
• D. The solution of MDP is to find a policy that maximizes the cumulative rewards
Q 1.1 Which of the following statement about MDP is **not** true?

- A. The reward function must output a scalar value (**True: need to be able to compare**)
- B. The policy maps states to actions (**True: a policy tells you what action to take for each state**).
- C. The probability of next state can depend on current and previous states (**False: Markov assumption**).
- D. The solution of MDP is to find a policy that maximizes the cumulative rewards (**True: want to maximize rewards overall**).
Q 2.1 Consider an MDP with 2 states \{A, B\} and 2 actions: “stay” at current state and “move” to other state. Let $r$ be the reward function such that $r(A) = 1$, $r(B) = 0$. Let $\gamma$ be the discounting factor. Let $\pi$: $\pi(A) = \pi(B) = \text{move}$ (i.e., an “always move” policy). What is the value function $V^\pi(A)$?

- A. 0
- B. $\frac{1}{1 - \gamma}$
- C. $\frac{1}{1 - \gamma^2}$
- D. 1
Q 2.1 Consider an MDP with 2 states \{A, B\} and 2 actions: “stay” at current state and “move” to other state. Let \( r \) be the reward function such that \( r(A) = 1, r(B) = 0 \). Let \( \gamma \) be the discounting factor. Let \( \pi: \pi(A) = \pi(B) = \text{move} \) (i.e., an “always move” policy). What is the value function \( V^\pi(A) \)?

- A. 0
- B. \( 1/(1-\gamma) \)
- C. \( 1/(1-\gamma^2) \)
- D. 1
Q 2.1 Consider an MDP with 2 states \{A, B\} and 2 actions: "stay" at current state and "move" to other state. Let \( r \) be the reward function such that \( r(A) = 1, \, r(B) = 0 \). Let \( \gamma \) be the discounting factor. Let \( \pi: \pi(A) = \pi(B) = \text{move} \) (i.e., an "always move" policy). What is the value function \( V^\pi(A) \)?

- A. 0
- B. \( 1/(1-\gamma) \)
- C. \( 1/(1-\gamma^2) \) (States: A,B,A,B,... rewards 1,0, \( \gamma^2 \),0, \( \gamma^4 \),0)
- D. 1