Q1-1: Select the correct statement.

A. *Markov Assumption implies that given the present state and action, all following states are independent of all past states.*

B. *All Reinforcement Learning techniques adopts Markov assumption property.*

1. Both the statements are TRUE.
2. Statement A is TRUE, but statement B is FALSE.
3. Statement A is FALSE, but statement B is TRUE.
4. Both the statements are FALSE.

Q1-1: Select the correct statement.

A. *Markov Assumption implies that given the present state and action, all following states are independent of all past states.*

B. *All Reinforcement Learning techniques adopts Markov assumption property.*

1. Both the statements are TRUE.
2. Statement A is TRUE, but statement B is FALSE. ⟵
3. Statement A is FALSE, but statement B is TRUE.
4. Both the statements are FALSE.

Though markov assumption makes the analysis easier, it's not necessary to assume Markov property.

Break & Quiz

Q 1.2 Consider an MDP with 2 states {*A, B*} and 2 actions: **"stay"** at current state and "**move**" to other state; suppose we are guaranteed to transition according to the action. Let **r** be the reward function such that **r**(*A*) = 1, **r**(*B*) = 0. Let γ be the discounting factor. What is the optimal policy $\pi(A)$ and $\pi(B)$? What are $V^*(A)$, $V^*(B)$?

- A. Stay, Stay, 1/(1-γ), 1
- B. Stay, Move, 1/(1-γ), 1/(1-γ)
- C. Move, Move, 1/(1-γ), 1
- D. Stay, Move, 1/(1-γ), γ/(1-γ)

Break & Quiz

Q 1.2 Consider an MDP with 2 states {*A, B*} and 2 actions: **"stay"** at current state and "**move**" to other state; suppose we are guaranteed to transition according to the action. Let **r** be the reward function such that **r**(*A*) = 1, **r**(*B*) = 0. Let γ be the discounting factor. What is the optimal policy $\pi(A)$ and $\pi(B)$? What are $V^*(A)$, $V^*(B)$?

- A. Stay, Stay, 1/(1-γ), 1
- B. Stay, Move, 1/(1-γ), 1/(1-γ)
- C. Move, Move, 1/(1-γ), 1
- **D. Stay, Move, 1/(1-γ), γ/(1-γ)**

Break & Quiz

Q 1.2 Consider an MDP with 2 states {*A, B*} and 2 actions: **"stay"** at current state and "**move**" to other state; suppose we are guaranteed to transition according to the action. Let **r** be the reward function such that **r**(*A*) = 1, **r**(*B*) = 0. Let γ be the discounting factor. What is the optimal policy $\pi(A)$ and $\pi(B)$? What are $V^*(A)$, $V^*(B)$?

- A. Stay, Stay, 1/(1-γ), 1
- B. Stay, Move, 1/(1-γ), 1/(1-γ)
- C. Move, Move, 1/(1-γ), 1
- **D. Stay, Move, 1/(1-γ), γ/(1-γ)** Note: want to stay at A, if at B, move to A. Starting at A, sequence A,A,A,… rewards 1, γ, γ^2,…. Start at B, sequence B,A,A,… rewards 0, γ, γ^2,…. Sums to 1/(1-γ), γ/(1-γ).

Break & Quiz

Q 2.1 For Q learning to converge to the true Q function, we must

- A. Visit every state and try every action
- B. Perform at least 20,000 iterations.
- C. Re-start with different random initial table values.
- D. Prioritize exploitation over exploration.

Break & Quiz

Q 2.1 For Q learning to converge to the true Q function, we must

- **A. Visit every state and try every action**
- B. Perform at least 20,000 iterations.
- C. Re-start with different random initial table values.
- D. Prioritize exploitation over exploration.

Break & Quiz

Q 2.1 For Q learning to converge to the true Q function, we must

- **A. Visit every state and try every action**
- B. Perform at least 20,000 iterations. (No: this is dependent on the particular problem, not a general constant).
- C. Re-start with different random initial table values. (No: this is not necessary in general).
- D. Prioritize exploitation over exploration. (No: insufficient exploration means potentially unupdated state action pairs).