

Q Search models, datasets, users...

 \equiv

Mid-way Quiz

The best way to learn and to avoid the illusion of competence is to test yourself. This will help you to find where you need to reinforce your knowledge.

Q1: What are the two main approaches to find optimal policy?

✓ Policy-based methods

Correct! With Policy-Based methods, we train the policy directly to learn which action to take given a state.

- Random-based methods
- ✓ Value-based methods

Correct! With value-based methods, we train a value function to learn which state is more valuable and use this value function to take the action that leads to it.

Evolution-strategies methods

Submit

You got all the answers!

Q2: What is the Bellman Equation?

▼ Solution

The Bellman equation is a recursive equation that works like this: instead of starting for each state from the beginning and calculating the return, we can consider the value of any state as:

Rt+1 + gamma * V(St+1)

The immediate reward + the discounted value of the state that follows

Q3: Define each part of the Bellman Equation

The Bellman Equation

$$V_{\pi}(s) = \mathbf{F}_{\pi}[R_{t+1} + \gamma * V_{\pi}(S_{t+1}) | S_t = s]$$

Deep RL Course documentation

Mid-way Quiz >



Q

▼ Solution

The Bellman Equation

$$V_{\pi}(s) = \mathbf{F}_{\pi}[R_{t+1} + \gamma * V_{\pi}(S_{t+1}) | S_t = s]$$

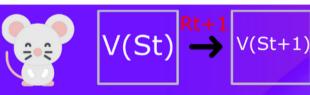
Value of state s

Expected value of immediate reward

+ the discounted value of next_state

If the agent starts at state s

And uses the policy tochoose its actions for all time steps



Deep RL Course

V(St) = Rt+1 + gamma * V(St+1)

Q4: What is the difference between Monte Carlo and Temporal Difference learning methods?

With Monte Carlo methods, we update the value function from a complete episode

Correct!

- With Monte Carlo methods, we update the value function from a step
- With TD learning methods, we update the value function from a complete episode
- With TD learning methods, we update the value function from a step

Correct!

Submit

You got all the answers!

Q5: Define each part of Temporal Difference learning formula

TD Learning Approach:

Temporal Difference Learning: learning at each time step.

$$V(S_t) \leftarrow V(S_t) + \alpha[R_{t+1} + \gamma V(S_{t+1}) - V(S_t)]$$

▼ Solution

TD Learning Approach:

Temporal Difference Learning: learning at each time step.

$$V(S_t) \leftarrow V(S_t) + lpha[R_{t+1} + \gamma V(S_{t+1}) - V(S_t)]$$

New value of state t

Former Learning Reward estimation of Rate value of state

Discounted value of next state

TD Target

Q6: Define each part of Monte Carlo learning formula

Monte Carlo Approach:

Monte Carlo: waits until the end of the episode, then calculates Gt (return) and uses it as a target for its value or policy.

$$V(S_t) \leftarrow V(S_t) + \alpha [G_t - V(S_t)]$$

▼ Solution

Monte Carlo Approach:

Monte Carlo: waits until the end of the episode, then calculates Gt (return) and uses it as a target for its value or policy.

$$V(S_t) \leftarrow V(S_t) + \alpha [G_t - V(S_t)]$$

New value of state t

Former estimation of value of state t (= Expected return starting at that state)

Learning Return at timestep

Former estimation of value of state t (= Expected return starting at that state)

Congrats on finishing this Quiz , if you missed some elements, take time to read again the previous sections to reinforce (your knowledge.

← Mid-way Recap

Introducing Q-Learning →