

Q Search models, datasets, users...



Second Quiz

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

Q1: What is Q-Learning?

The algorithm we use to train our Q-function

Correct!

- A value function
- An algorithm that determines the value of being at a particular state and taking a specific action at that state

Correct!

A table

Submit

You got all the answers!

Q2: What is a Q-table?

- An algorithm we use in Q-Learning
- Q-table is the internal memory of our agent

Correct!

In Q-table each cell corresponds a state value

Submit

You got all the answers!

Q3: Why if we have an optimal Q-function Q* we have an optimal policy?

▼ Solution

Because if we have an optimal Q-function, we have an optimal policy since we know for each state what is the best action to take.

The link between Value and Policy:

$$\pi^*(s) = rg \max_a Q^*(s,a)$$

Finding an optimal value function leads to having an optimal policy.

Q4: Can you explain what is Epsilon-Greedy Strategy?

▼ Solution

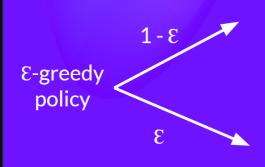
Epsilon Greedy Strategy is a policy that handles the exploration/exploitation trade-off.

The idea is that we define epsilon $\varepsilon = 1.0$:

- With probability 1ε : we do exploitation (aka our agent selects the action with the highest state-action pair value).
- With probability ε : we do exploration (trying random action).

Q-Learning, Step 2

Choose action A_t using policy derived from Q (e.g., ϵ -greedy)



Exploitation (selects the greedy action)

Exploration (selects a random action)

Choose the action using E-greedy policy

Q5: How do we update the Q value of a state, action pair?

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha[R_{t+1} + \gamma max_aQ(S_{t+1}, a) - Q(S_t, A_t)]$$

▼ Solution

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha[R_{t+1} + \gamma max_aQ(S_{t+1}, a) - Q(S_t, A_t)]$$

New Q-value estimation

Q-value estimation Rate Reward

Former Learning Immediate Discounted Estimate optimal Q-value of next state

Former Q-value estimation

Deep RL Course documentation

Second Quiz ~



TD Error

Q6: What's the difference between on-policy and off-policy

▼ Solution

Off-policy vs On-policy

Off-policy: using a different policy for acting and for updating.

Choose action A_t using policy derived from Q (e.g., ϵ -greedy) Epsilon Greedy Take action A_t and observe R_{t+1}, S_{t+1} $Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha(R_{t+1} + \gamma \max_a Q(S_{t+1}, a) - Q(S_t, A_t))$

On-policy: using the same policy for acting and updating.

Choose action A_0 using policy derived from Q (e.g., ϵ -greedy) $t \leftarrow 0$ Epsilon Greedy Policy repeat Take action A_t and observe R_{t+1}, S_{t+1} Choose action A_{t+1} using policy derived from Q (e.g., ϵ -greedy) $Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha(R_{t+1} + \gamma Q(S_{t+1}, A_{t+1}))$

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

← Hands-on	Conclusion	\rightarrow