Search models, datasets, users...

Second 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 is Q-Learning?

- [x] The algorithm we use to train our Q-function

Correct!

- [ ] A value function
- [x] 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
- [x] Q-table is the internal memory of our agent

Correct!

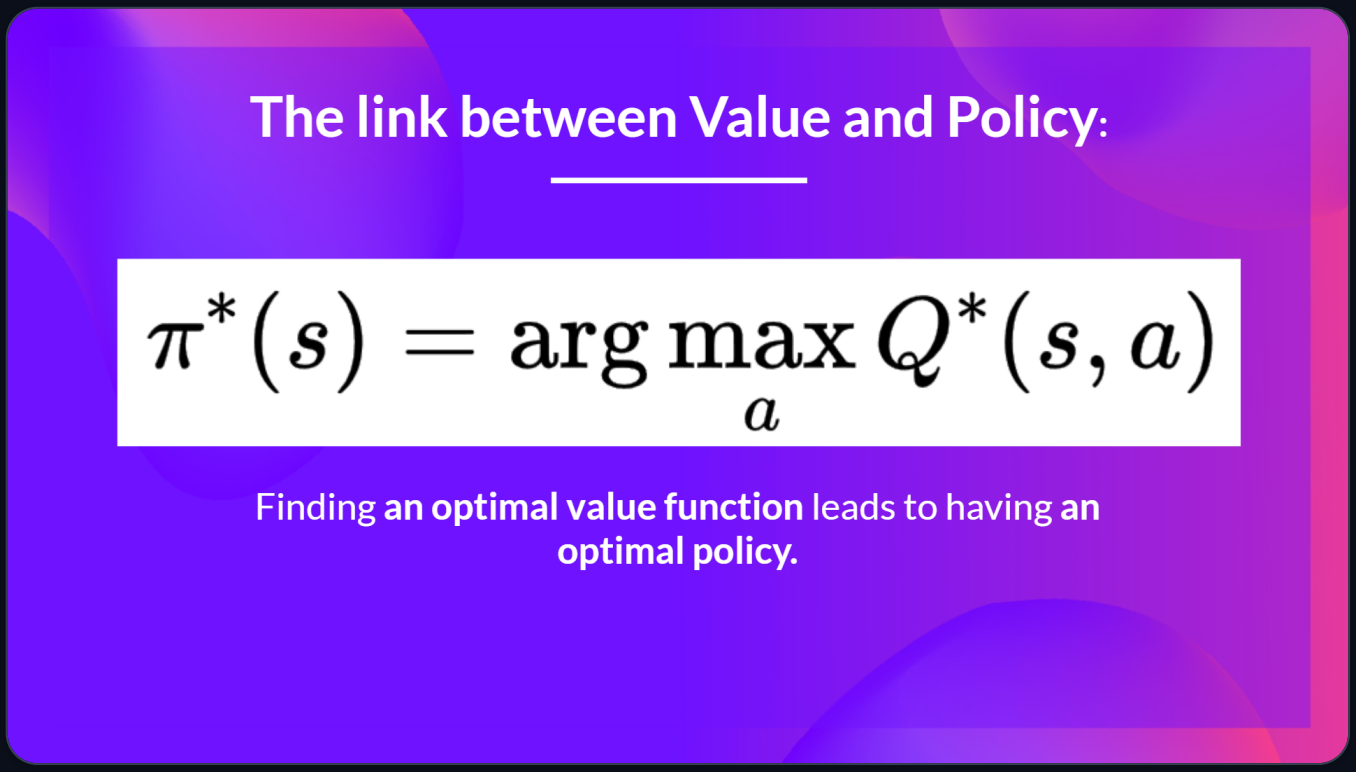
- [ ] In Q-table each cell corresponds a state value

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.



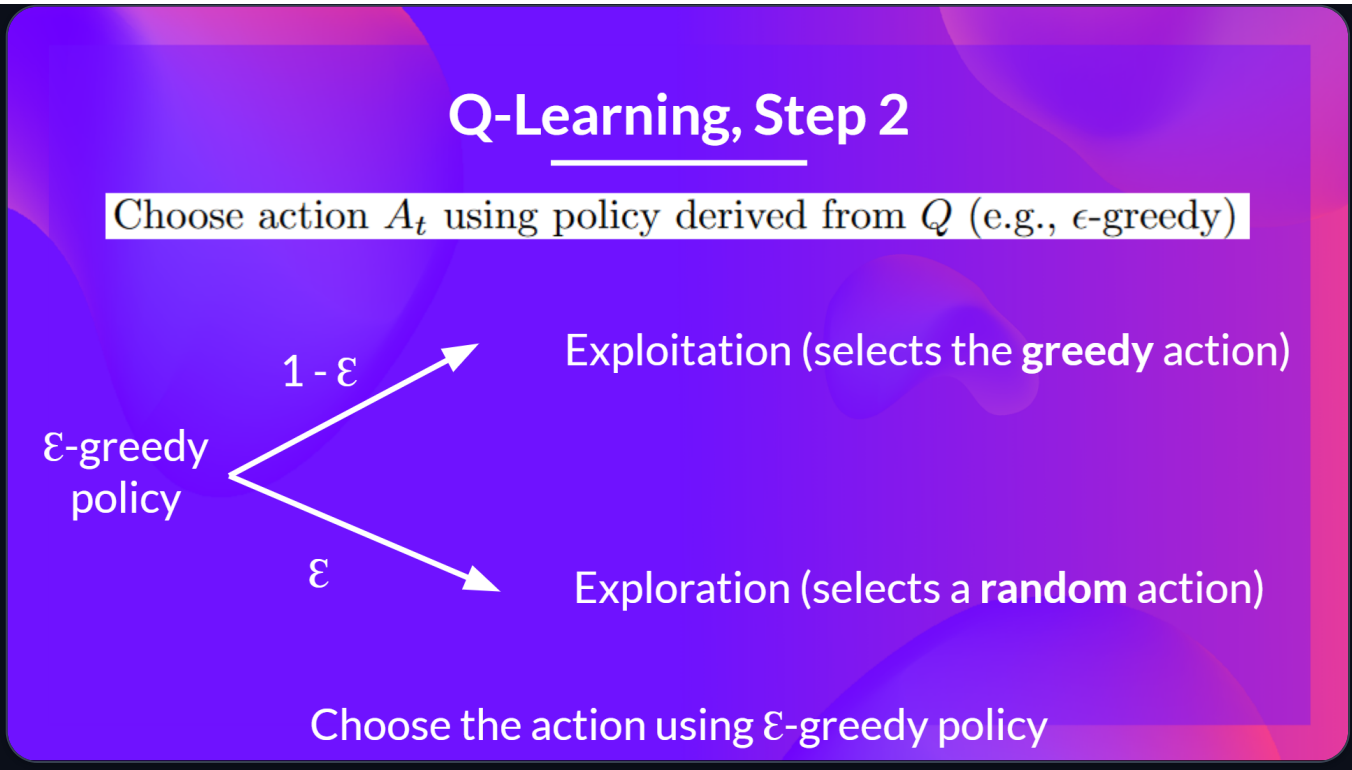
Q4: Can you explain what is Epsilon-Greedy Strategy?

▼ Solution

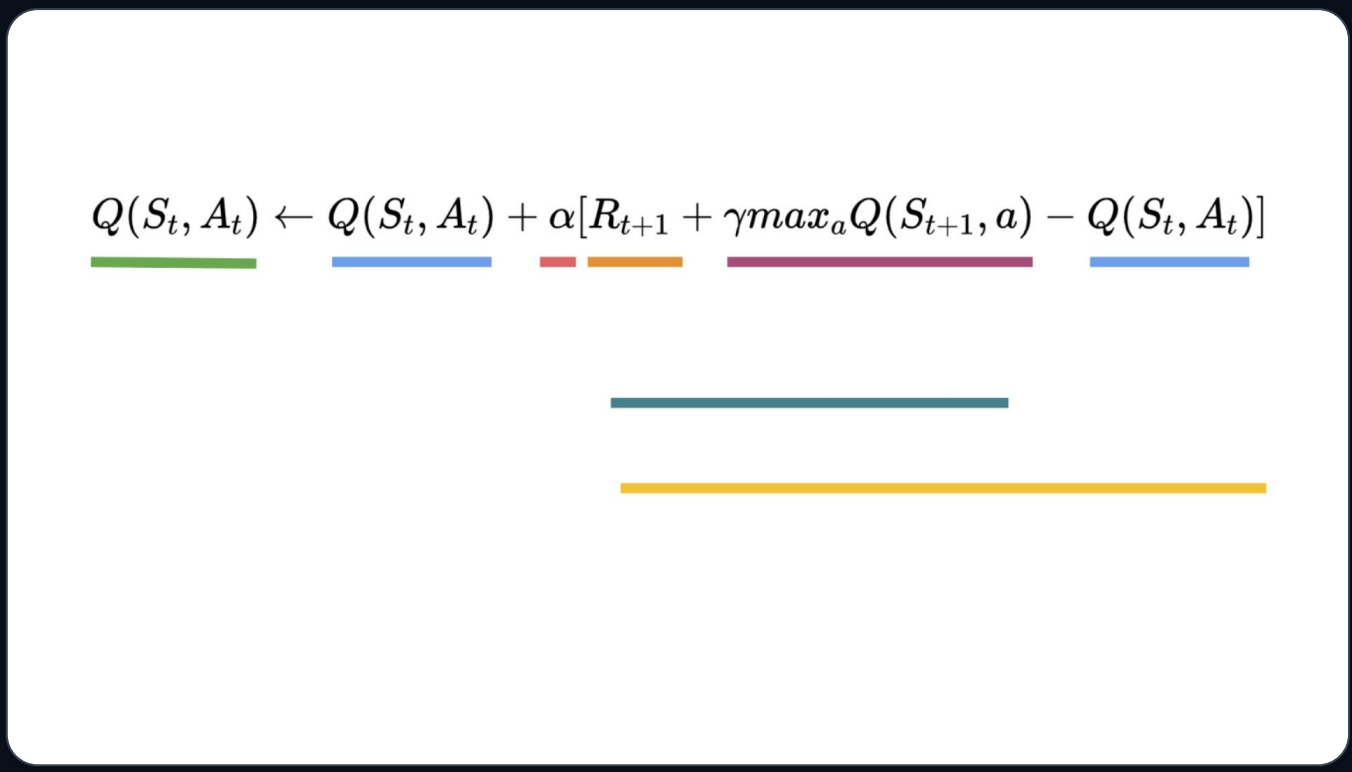
Epsilon Greedy Strategy is a policy that handles the exploration/exploitation trade-off.

The idea is that we define epsilon ε = 1.0:

- With *probability 1 — ε* : 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).



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



▼ Solution

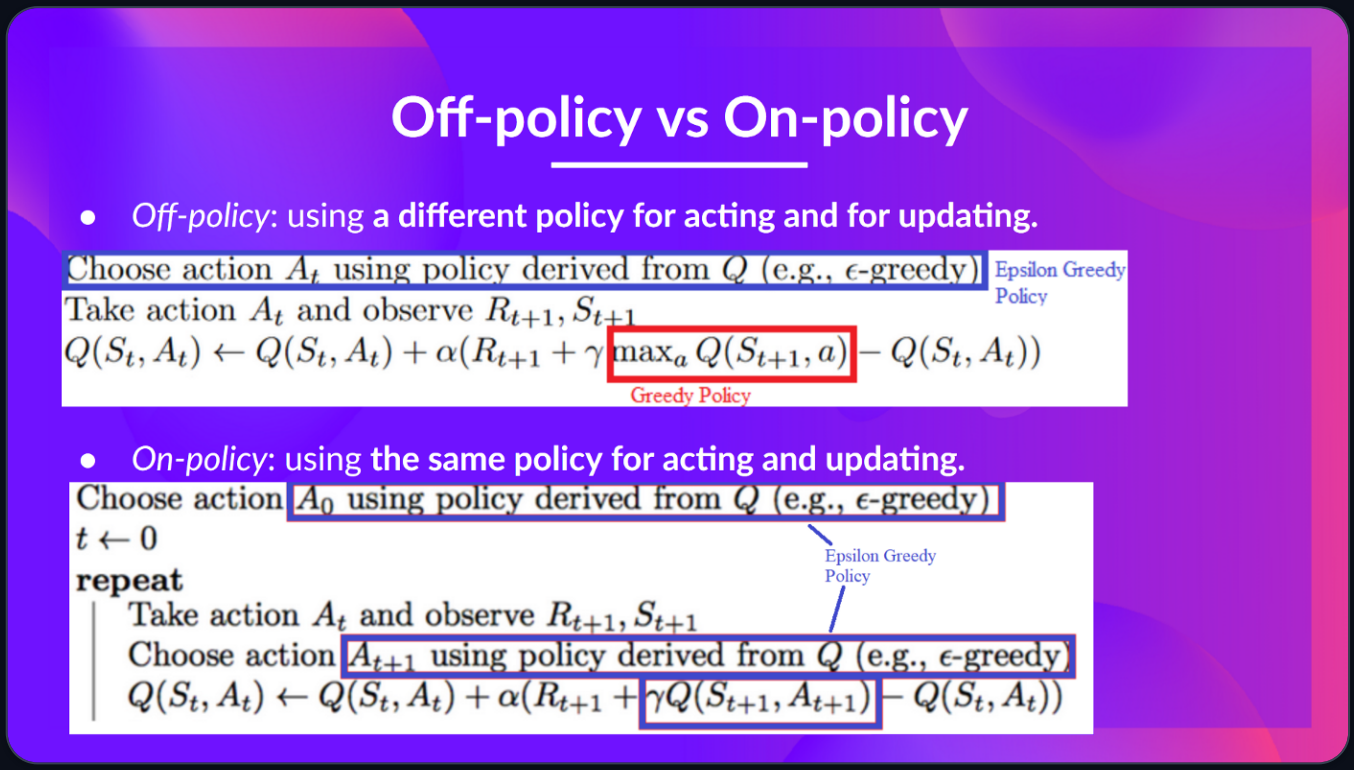
$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha[R_{t+1} + \gamma max_a Q(S_{t+1}, a) - Q(S_t, A_t)]$$

New Q-value estimation | Former Q-value estimation | Learning Rate | Immediate Reward | Discounted Estimate optimal Q-value of next state | Former Q-value estimation

TD Error

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

▼ Solution



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

← Hands-on Conclusion →