Поздравляем! Вы прошли тест!
ДЛЯ УСПЕШНОГО ПРОХОЖДЕНИЯ 80% или выше

Продолжить обучение

ОЦЕНКА
88.88%

Practice Quiz

ОБЩИЙ БАЛЛ 9

1. What is the target policy in Q-learning? 1 / 1 балл

- ϵ-greedy with respect to the current action-value estimates
- Greedy with respect to the current action-value estimates

Correct
Correct! Q-learning's target policy is greedy with respect to the current action-value estimates.

2. Which Bellman equation is the basis for the Q-learning update? 1 / 1 балл

- Bellman equation for state values
- Bellman equation for action values
- Bellman optimality equation for state values
- Bellman optimality equation for action values

Correct
Correct! The Q-learning update is based on the Bellman optimality equation for action values.

3. Which Bellman equation is the basis for the Sarsa update? 0 / 1 балл

- Bellman equation for state values
- Bellman equation for action values
- Bellman optimality equation for state values
- Bellman optimality equation for action values

Incorrect
Incorrect. Please review Lesson 1 (Video: GPI with TD: Sarsa)

4. Which Bellman equation is the basis for the Expected Sarsa update? 1 / 1 балл

- Bellman equation for state values
- Bellman equation for action values
- Bellman optimality equation for state values
- Bellman optimality equation for action values

Correct
Correct! The Expected Sarsa update is based on the Bellman equation for action values.

5. Which algorithm's update requires more computation per step? 1 / 1 балл

- Expected Sarsa
- Sarsa

Correct
Correct! Expected Sarsa computes the expectation over next actions.

6. Which algorithm has a higher variance target? 1 / 1 балл

- Expected Sarsa
- Sarsa

Correct
Correct! We saw that Sarsa was more sensitive to the choice of step-size because its target has higher variance.

7. Q-learning does not learn about the outcomes of exploratory actions. 1 / 1 балл

- True
- False

Correct

Correct! The update in Q-learning only learns about the greedy action. As demonstrated in Cliff World, it ignores the outcomes of exploratory actions.

8. **Sarsa, Q-learning, and Expected Sarsa have similar targets on a transition to a terminal state.** 1 / 1 балл

- (•) True
- ( ) False

✓ **Correct**
Correct! The target in this case only depends on the reward.

9. **Sarsa needs to wait until the end of an episode before performing its update.** 1 / 1 балл

- ( ) True
- (•) False

✓ **Correct**
Correct! Unlike Monte Carlo methods, Sarsa performs its updates at every time-step using the reward and the next action-value estimate.