Practice Quiz: Model-free reinforcement learning -- Q-learning and SARSA (not assessed)

Due No due date   **Points** 7   **Questions** 5   **Time Limit** None
Allowed Attempts Unlimited

Take the Quiz Again

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	less than 1 minute	0 out of 7

Submitted Jun 6 at 12:59

Unanswered

Question 1   0 / 1 pts

Model-free reinforcement learning is:

○ Q-learning but not SARSA

○ Any learning without a model

○ Learning the environment from experience and rewards so we can construct a policy

Correct Answer   ○ Learning a policy directly from experience and rewards

Unanswered   ### Question 2   0 / 1 pts

Which of the following is not true?

○ Q-learning learns Q values but SARSA does not

○ Q-learning is model-free learning

○ SARSA is model-free but can be used with a model

○ SARSA is on-policy learning

Question 3

0 / 1 pts

SARSA is an on-policy learning approach because:

○ It updates the Q values based on the actual policy followed

○ It updates based on the maximum estimated future reward, which is what we follow on the policy

○ It updates the policy directly

Question 4

0 / 2 pts

State	Action			
	North	South	East	West
(0,0)	0.53	0.36	0.36	0.21
(0,1)	0.61	0.27	0.23	0.23
. . .				
(2,2)	0.79	0.72	0.90	0.72
(2,3)	0.90	0.78	0.99	0.81

Assume a learning rate $\alpha = 0.1$ and discount factor $\gamma = 0.9$, update the Q value if the action North is chosen in state (0,0), receiving no reward, and transitioning to state (0,1)

0.5319 (with margin: 0.0019)

$$
\begin{aligned}
Q((0,0), N) \quad &\leftarrow \quad Q((0,0), N) + \alpha[r + \gamma \max_{a'} Q((0,0), a') - Q((0,0), N)] \\
&\leftarrow \quad 0.53 + 0.1[0 + 0.9 \cdot Q((0,1), N) - Q((0,0), N)] \\
&\leftarrow \quad 0.53 + 0.1[0 + 0.9 \cdot 0.61 - 0.53] \\
&\leftarrow \quad 0.5319
\end{aligned}
$$

Question 5   **0 / 2 pts**

State	Action			
	North	South	East	West
(0,0)	0.53	0.36	0.36	0.21
(0,1)	0.61	0.27	0.23	0.23
. . .				
(2,2)	0.79	0.72	0.90	0.72
(2,3)	0.90	0.78	0.99	0.81

Assume a learning rate $\alpha = 0.2$ and discount factor $\gamma = 0.9$, update the Q value if the action North is chosen in state (0,0), receiving no reward, and transitioning to state (0,1), and action South being chosen from state (0,1).

Note the learning rate parameter is 0.2, not 0.1 as in the previous example.

0.4726 (with margin: 0.0026)

$$
\begin{aligned}
Q((0,0), N) \quad &\leftarrow \quad Q((0,0), N) + \alpha[r + \gamma Q((0,1), S) - Q((0,0), N)] \\
&\leftarrow \quad 0.53 + 0.2[0 + 0.9 \cdot Q((0,1), S) - Q((0,0), N)] \\
&\leftarrow \quad 0.53 + 0.2[0 + 0.9 \cdot 0.27 - 0.53] \\
&\leftarrow \quad 0.4726
\end{aligned}
$$