Unit 11 - Week 9

Assignment 9

The due date for submitting this assignment has passed.
As per our records you have not submitted this assignment.

Due on 2020-04-01, 23:59 IST.

1) **Statement:** DQN is implemented with current and target network. *1 point*
Reason: Using target network helps in avoiding chasing a non-stationary target

- Both Assertion and Reason are true, and Reason is correct explanation for Assertion
- Both Assertion and Reason are true, but Reason is not correct explanation for assertion
- Assertion is true, Reason is false
- Both Assertion and Reason are false

No, the answer is incorrect.
Score: 0
Accepted Answers:
Both Assertion and Reason are true, and Reason is correct explanation for Assertion

2) Which of the following is true about DQN? *1 point*

- It may converge to non-optimal policy
- It is an off-policy technique
- It can be efficiently used for very large state spaces
- It can be efficiently used for continuous action spaces

No, the answer is incorrect.
Score: 0
Accepted Answers:
It may converge to non-optimal policy
It is an off-policy technique
It can be efficiently used for very large state spaces

3) Policy gradient methods can be used for continuous action spaces *1 point*

- True
- False

No, the answer is incorrect.
Score: 0
Accepted Answers:
True

4) **Assertion:** Actor-critic updates have lesser variance than REINFORCE updates *1 point*
Reason: Actor-critic methods use TD target instead of G_t

- Both Assertion and Reason are true, and Reason is correct explanation for Assertion
- Both Assertion and Reason are true, but Reason is not correct explanation for assertion
- Assertion is true, Reason is false
- Both Assertion and Reason are false

No, the answer is incorrect.
Score: 0
Accepted Answers:
Both Assertion and Reason are true, and Reason is correct explanation for Assertion

5) Policy Gradient Theorem does not hold for average reward formulation *1 point*

- True
- False

No, the answer is incorrect.
Score: 0
Accepted Answers:
False

6) Why is an experience replay buffer used instead of direct updates to network in DQN? *1 point*

- Random sampling from experience replay buffer breaks correlations among transitions
- If not, the off-policy nature of Q-Learning is lost
- It guarantees convergence to the optimal policy
- None of the above

No, the answer is incorrect.
Score: 0
Accepted Answers:
Random sampling from experience replay buffer breaks correlations among transitions

7) Which of the following is the correct definition of average reward formulation? *1 point*

- $\rho(\pi) = \lim_{N\to\infty} \mathbb{E}[r_1 + r_2 + \ldots + r_N]$
- $\rho(\pi) = \lim_{N\to\infty} \mathbb{E}[r_1 + r_2 + \ldots + r_N | \pi]$
- $\rho(\pi) = \lim_{N\to\infty} \frac{1}{N}\mathbb{E}[r_1 + r_2 + \ldots + r_N]$
- $\rho(\pi) = \lim_{N\to\infty} \frac{1}{N}\mathbb{E}[r_1 + r_2 + \ldots + r_N | \pi]$

No, the answer is incorrect.
Score: 0
Accepted Answers:
$\rho(\pi) = \lim_{N\to\infty} \frac{1}{N}\mathbb{E}[r_1 + r_2 + \ldots + r_N | \pi]$

8) Choose the correct statement for Policy Gradient Theorem for average reward formulation: *1 point*

- $\frac{\partial\rho(\pi)}{\partial\theta} = \sum_s d^\pi(s) \sum_a \frac{\partial\pi(s,a)}{\partial\theta}$
- $\frac{\partial\rho(\pi)}{\partial\theta} = \sum_s v^\pi(s) \sum_a \frac{\partial\pi(s,a)}{\partial\theta} q^\pi(s,a)$
- $\frac{\partial\rho(\pi)}{\partial\theta} = \sum_s d^\pi(s) \sum_a \frac{\partial\pi(s,a)}{\partial\theta} q^\pi(s,a)$
- None of the above

No, the answer is incorrect.
Score: 0
Accepted Answers:
$\frac{\partial\rho(\pi)}{\partial\theta} = \sum_s d^\pi(s) \sum_a \frac{\partial\pi(s,a)}{\partial\theta} q^\pi(s,a)$

9) In Actor-critic algorithm, suppose that Q^π is approximated and the approximation is compatible with the parameterization of the actor. Assuming differentiable function approximators, which of the following can be concluded? *1 point*

- Convergence to globally optimal policy
- Convergence to locally optimal policy
- Cannot comment on the convergence
- Does not converge at all

No, the answer is incorrect.
Score: 0
Accepted Answers:
Convergence to locally optimal policy

10) Using similar parameterizations to represent policies, Monte Carlo policy gradient methods would converge to a locally optimal policy faster than Actor critic method in which approximation in critic is compatible with actor parameterization? *1 point*

- True
- False

No, the answer is incorrect.
Score: 0
Accepted Answers:
False