

Policy Gradient Methods

← Back Practice Assignment • 45 min

🌐 English ▾

Your grade: 95.83%

Your latest: 95.83% • Your highest: 95.83%
To pass you need at least 80%. We keep your highest score.

Next item →

1. Which of the following is true about policy gradient methods? (Select all that apply) 1 / 1 point

☒ If we have access to the true value function v_π , we can perform unbiased stochastic gradient updates using the result from the Policy Gradient Theorem.

☒ **Correct**
Correct. We derived this stochastic update by multiplying and dividing by $\pi(A|S)$.

☒ The policy gradient theorem provides a form for the policy gradient that does not contain the gradient of the state distribution μ , which is hard to estimate.

☒ **Correct**
Correct.

☒ Policy gradient methods do gradient ascent on the policy objective.

☒ **Correct**
Correct. Policy gradient methods maximize the policy objective, and hence perform gradient ascent.

☐ Policy gradient methods use generalized policy iteration to learn policies directly.

2. Which of the following statements about parameterized policies are true? (Select all that apply) 1 / 1 point

☐ The function used for representing the policy must be a softmax function.

☒ The probability of selecting any action must be greater than or equal to zero.

☒ **Correct**
Correct! This is one of the conditions for a valid probability distribution.

☒ For each state, the sum of all the action probabilities must equal to one.

☒ **Correct**
Correct! This condition is necessary for the function to be a valid probability distribution.

☐ The policy must be approximated using linear function approximation.

3. Assume you're given the following preferences $h_1 = 44$, $h_2 = 42$, and $h_3 = 38$, corresponding to three different actions (a_1, a_2, a_3), respectively. Under a softmax policy, what is the probability of choosing a_2 , rounded to three decimal numbers? 1 / 1 point

☐ 0.42

☒ 0.119

☐ 0.879

☐ 0.002