Progress

Mentor

1 point

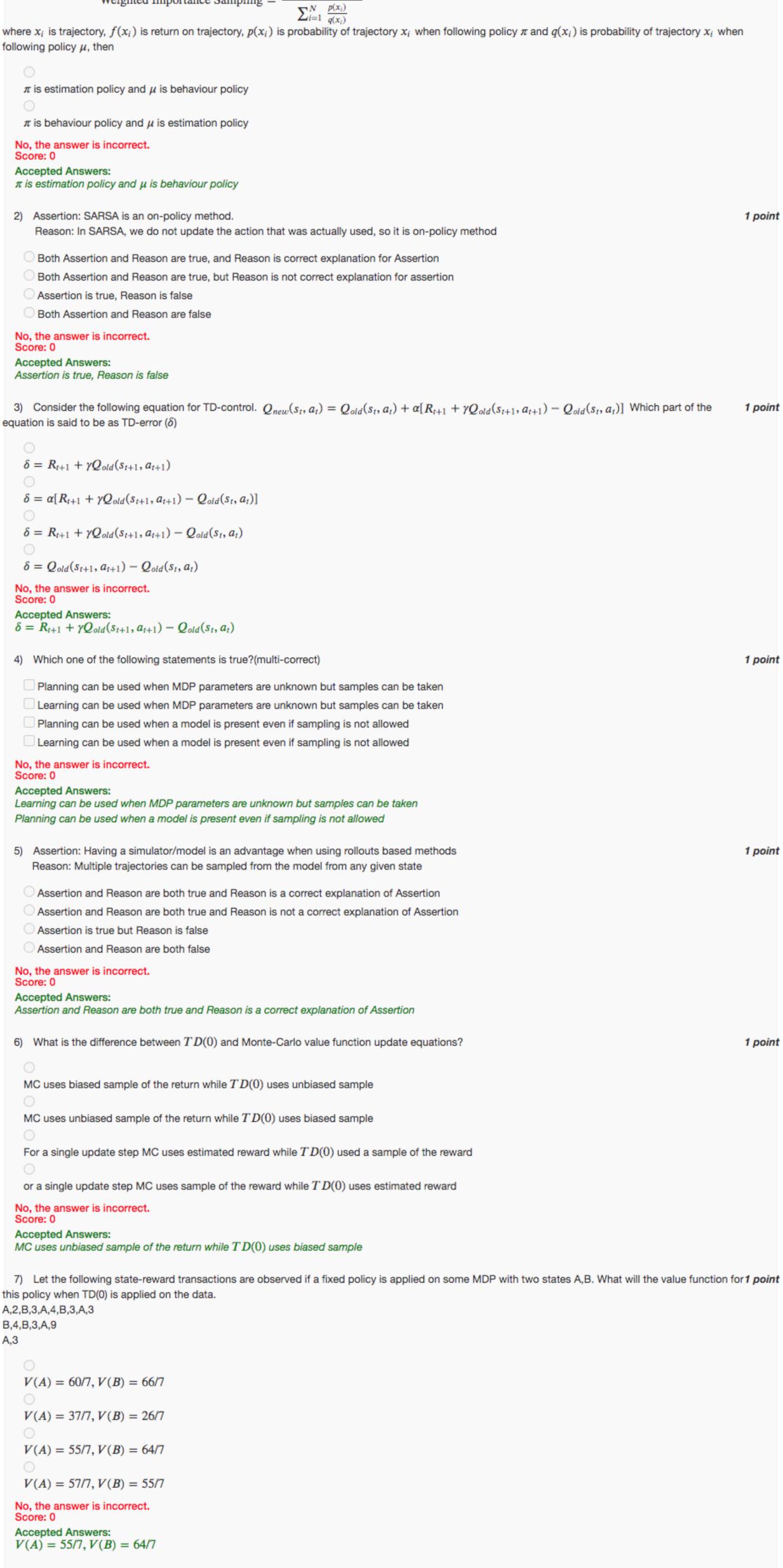
1 point

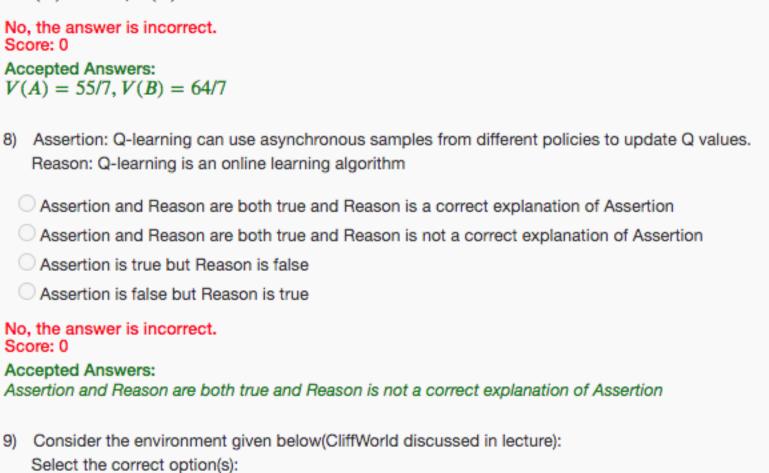
1 point

1 point

Unit 8 - Week 6

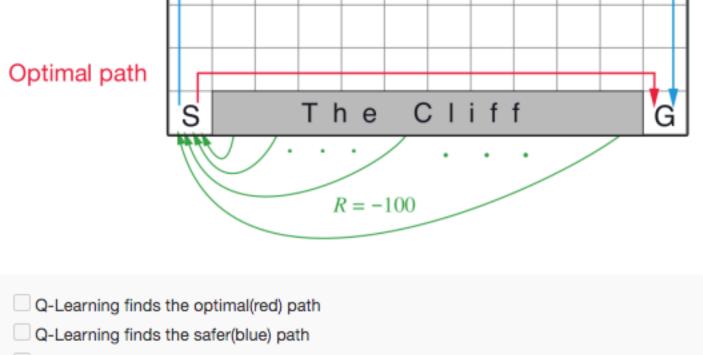
Course outline Assignment 6 How does an NPTEL online The due date for submitting this assignment has passed. Due on 2020-03-11, 23:59 IST. course work? As per our records you have not submitted this assignment. Week 0 Consider the following equation for weighted Importance Sampling Weighted Importance Sampling = $\frac{\sum_{i=1}^{N} f(x_i) \frac{p(x_i)}{q(x_i)}}{\sum_{i=1}^{N} \frac{p(x_i)}{q(x_i)}}$ Week 1 Week 2 following policy μ , then Week 3 Week 4 π is estimation policy and μ is behaviour policy Week 5 π is behaviour policy and μ is estimation policy Week 6 No, the answer is incorrect. Score: 0 Off Policy MC Accepted Answers: π is estimation policy and μ is behaviour policy OUCT Assertion: SARSA is an on-policy method. TD(0) Reason: In SARSA, we do not update the action that was actually used, so it is on-policy method TD(0) Control Both Assertion and Reason are true, and Reason is correct explanation for Assertion Q-Learning Both Assertion and Reason are true, but Reason is not correct explanation for assertion. Afterstate Assertion is true, Reason is false Quiz : Assignment 6 Both Assertion and Reason are false No, the answer is incorrect. Week 7 Score: 0 Accepted Answers: Week 8 Assertion is true, Reason is false Week 9 equation is said to be as TD-error (δ) Week 10 $\delta = R_{t+1} + \gamma Q_{old}(s_{t+1}, a_{t+1})$ Week 11 $\delta = \alpha [R_{t+1} + \gamma Q_{old}(s_{t+1}, a_{t+1}) - Q_{old}(s_t, a_t)]$ Week 12 $\delta = R_{t+1} + \gamma Q_{old}(s_{t+1}, a_{t+1}) - Q_{old}(s_t, a_t)$ DOWNLOAD VIDEOS Assignment Solutions $\delta = Q_{old}(s_{t+1}, a_{t+1}) - Q_{old}(s_t, a_t)$ No, the answer is incorrect. NPTEL Resources Score: 0 Accepted Answers: $\delta = R_{t+1} + \gamma Q_{old}(s_{t+1}, a_{t+1}) - Q_{old}(s_t, a_t)$ Which one of the following statements is true?(multi-correct) Planning can be used when MDP parameters are unknown but samples can be taken Learning can be used when MDP parameters are unknown but samples can be taken Planning can be used when a model is present even if sampling is not allowed Learning can be used when a model is present even if sampling is not allowed No, the answer is incorrect. Score: 0 Accepted Answers: Learning can be used when MDP parameters are unknown but samples can be taken Planning can be used when a model is present even if sampling is not allowed



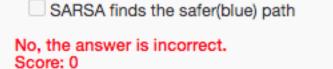


Safer path Optimal path G The

R = -1



It reduces the number of estimates if lot of state-action pairs lead to same afterstate



Q-Learning finds the optimal(red) path

SARSA finds the optimal(red) path

SARSA finds the safer(blue) path

Accepted Answers:

A,3

10) Which of the following is a benefit of keeping value estimates of afterstates instead of state- action pairs?

Both (a) and (b) None of the above

We can use it for lookahead in deterministic state transition MDPs

No, the answer is incorrect.

Score: 0 Accepted Answers:

Both (a) and (b)