Unit 7 - Week 5

Assignment 5

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

Due on 2020-03-04, 23:59 IST.

1) Which of the following equation state that L is a contraction on value function? *1 point*

- $\exists s,\ |Lv(s) - Lu(s)| \geq \gamma||v - u||$
- $\forall s,\ |Lv(s) - Lu(s)| \geq \gamma||v - u||$
- $\exists s,\ |Lv(s) - Lu(s)| \leq \gamma||v - u||$
- $\forall s,\ |Lv(s) - Lu(s)| \leq \gamma||v - u||$

No, the answer is incorrect.
Score: 0

Accepted Answers:
$\forall s,\ |Lv(s) - Lu(s)| \leq \gamma||v - u||$

2) In the value iteration algorithm, the stopping condition is given as follows: *1 point*

$$\textbf{if}\quad ||v^{n+1} - v^n|| < \frac{\epsilon(1-\gamma)}{2\gamma}:$$
$$\text{terminate}$$

What guarantee does such a stopping condition provide?

- The final policy obtained will be the optimal policy, π^*
- The final value estimates will be ϵ-close to optimal values, v^*
- The final value estimates will be optimal values, v^*
- The value estimates will be γ-close to optimal values, v^*

No, the answer is incorrect.
Score: 0

Accepted Answers:
The final value estimates will be ϵ-close to optimal values, v^*

3) If we apply the value iteration algorithm, at the end of step 4 we get a policy π . Is v^{n+1} the fixed point of π *1 point*

- yes
- no

No, the answer is incorrect.
Score: 0

Accepted Answers:
no

4) For both value and policy iteration algorithms we will get a sequence of vectors after some iterations, say $v_1, v_2 \ldots v_n$ for value iteration and $v_1', v_2' \ldots v_n'$ for policy iteration. Which of the following statements are true. *1 point*

- For all $v_i \in v_1, v_2 \ldots v_n$ there exists a policy for which v_i is a fixed point
- For all $v_i' \in v_1', v_2' \ldots v_n'$ there exists a policy for which v_i' is a fixed point.
- For all $v_i \in v_1, v_2 \ldots v_n$ there may not exist a policy for which v_i is a fixed point
- For all $v_i' \in v_1', v_2' \ldots v_n'$ there may not exist a policy for which v_i' is a fixed point

No, the answer is incorrect.
Score: 0

Accepted Answers:
For all $v_i' \in v_1', v_2' \ldots v_n'$ there exists a policy for which v_i' is a fixed point.
For all $v_i \in v_1, v_2 \ldots v_n$ there may not exist a policy for which v_i is a fixed point

5) If v satisfy $Lv = L_\pi v = v$ then *1 point*

- π is ϵ optimal and v is the fixed point of π
- π is not optimal and v is not the fixed point of π
- π is an optimal policy and v is the fixed point of π
- π is not optimal and v is the fixed point of the optimal policy

No, the answer is incorrect.
Score: 0

Accepted Answers:
π is an optimal policy and v is the fixed point of π

6) In some dynamic programming approaches, instead of updating all the states for large number of times, it is sometimes preferred to update some states by sampling randomly or selecting states seen in a trajectory. Which is the best reason for that? *1 point*

- Such methods(updating only subset of states at a time) have shown better convergence guarantees
- Many states are seldom visited so there is actually no need to update them at all
- Generally there are large number of states, and due to limited computation power, up- dating all states everytime is not affordable
- None of the above

No, the answer is incorrect.
Score: 0

Accepted Answers:
Generally there are large number of states, and due to limited computation power, up- dating all states everytime is not affordable

7) **Assertion:** If a single state's value function is updated between successive policy improvement, then the state's value function will converge to optimal value function, given that all states are updated. *1 point*
Reason: Generalized Policy Iteration(GPI) states that repeated interaction of policy evaluation and policy improvement process(under condition that all states are updated) leads to optimality, independent of other details.

- 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
- Assertion is false

No, the answer is incorrect.
Score: 0

Accepted Answers:
Both assertion and reason are true, and Reason is correct explanation for assertion

8) What is meant by off-policy Monte Carlo value function approximation? *1 point*

- The policy being evaluated is the same as the policy used to generate samples
- The policy being evaluated is different from the policy used to generate samples
- The policy being learnt is different from the policy used to generate samples

No, the answer is incorrect.
Score: 0

Accepted Answers:
The policy being evaluated is different from the policy used to generate samples

9) **Assertion:** Monte Carlo value function approximation methods need knowledge of model to be implemented. *1 point*
Reason: Monte Carlo value function approximation methods require a way to sample trajectories from the environment.

- Assertion and Reason are both true and Reason is a correct explanation of Assertion
- Assertion and Reason are both true and Reason is not a correct explanation of Assertion.
- Assertion is true but Reason is false
- Assertion is false but Reason is true

No, the answer is incorrect.
Score: 0

Accepted Answers:
Assertion is false but Reason is true

10) In every visit Monte Carlo methods, multiple samples are obtained from a single trajectory. *1 point*
Is it true that this leads to an increase in variance of the estimate?

- True
- False

No, the answer is incorrect.
Score: 0

Accepted Answers:
False