Unit 6 - Week 4

Assignment 4

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

Due on 2020-02-26, 23:59 IST.

1) Which of the following is a benefit of using RL algorithms for solving MDPs? *1 point*

- They do not require the state of the agent for solving MDP
- They do not require the action taken by the agent for solving MDP
- They do not require state transition probability for solving MDP
- hey do not require reward signal for solving MDP

No, the answer is incorrect.
Score: 0
Accepted Answers:
They do not require state transition probability for solving MDP

2) Select all the correct option(s) from the following equations: *1 point*

- $v^{\pi}(s) = \mathbb{E}_{\pi}[\sum_{i=t}^{T} \gamma^{i-t} R_{i+1} | S_t = s]$
- $q^{\pi}(s,a) = \sum_{s'} p(s'|s,a) v^{\pi}(s')$
- $v^{\pi}(s) = \sum_{a} \pi(a|s) q^{\pi}(s,a)$
- $q^{\pi}(s,a) = \sum_{s'} p(s'|s,a)[\mathbb{E}[r|s,a,s'] + \gamma \sum_{a'} \pi(a'|s') q^{\pi}(s',a')]$

No, the answer is incorrect.
Score: 0
Accepted Answers:
$v^{\pi}(s) = \mathbb{E}_{\pi}[\sum_{i=t}^{T} \gamma^{i-t} R_{i+1} | S_t = s]$
$q^{\pi}(s,a) = \sum_{s'} p(s'|s,a) v^{\pi}(s')$
$v^{\pi}(s) = \sum_{a} \pi(a|s) q^{\pi}(s,a)$
$q^{\pi}(s,a) = \sum_{s'} p(s'|s,a)[\mathbb{E}[r|s,a,s'] + \gamma \sum_{a'} \pi(a'|s') q^{\pi}(s',a')]$

3) Select the correct Bellman optimality equation: *1 point*

- $v^{*}(s) = \max_{a} \sum_{s'} p(s'|s,a)[\mathbb{E}[r|s,a,s'] + \gamma v^{*}(s')]$
- $v^{*}(s) = \max_{a} \sum_{s'} p(s'|s,a) v^{*}(s')$
- $v^{*}(s) = \max_{a} \sum_{s'} p(s'|s,a)[\gamma \mathbb{E}[r|s,a,s'] + v^{*}(s')]$
- $v^{*}(s) = \max_{a} \sum_{s'} p(s'|s,a) \gamma [\mathbb{E}[r|s,a,s'] + v^{*}(s')]$

No, the answer is incorrect.
Score: 0
Accepted Answers:
$v^{*}(s) = \max_{a} \sum_{s'} p(s'|s,a)[\mathbb{E}[r|s,a,s'] + \gamma v^{*}(s')]$

4) State True/False for the following statement. *1 point*
While solving MDPs, in case of discounted rewards, the value of γ(discount factor) can change the speed of learning but the final policy that would be learnt remains the same.

- True
- False

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

5) State True/False for the following statement *1 point*
In MDPs, there is a unique resultant state for any given state-action pair

- True
- False

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

6) Which of the following should be known for modelling an MDP? *1 point*
[i] Complete set of States
[ii] Set of action values
[iii] transition probabilities for state and actions
[iv] reward values for every allowed transition between states

- i,ii,iii, iv
- i, ii
- i,ii,iii
- iii, iv

No, the answer is incorrect.
Score: 0
Accepted Answers:
i,ii,iii, iv

7) Which of the following statements are true for an MDP? *1 point*

- An MDP with stochastic rewards may not have a deterministic optimal policy
- There can be multiple stochastic policies which are as good as the optimal deterministic policy
- If $0 \leq \gamma < 1$, then rank of the matrix $I - \gamma P_{\pi}$ is equal to $|S|$
- If $0 \leq \gamma < 1$, then rank of the matrix $I - \gamma P_{\pi}$ is less than $|S|$

No, the answer is incorrect.
Score: 0
Accepted Answers:
There can be multiple stochastic policies which are as good as the optimal deterministic policy
If $0 \leq \gamma < 1$, then rank of the matrix $I - \gamma P_{\pi}$ is equal to $|S|$

8) Consider an MDP with 3 states A,B,C. At each state we can go to either of the two states. i.e if we are in state A then we can perform 2 actions, going to state B or C. The rewards for each transactions are $r(A,B) = 3$ (reward if we go from A to B), $r(B,A) = 4$, $r(B,C) = 4$, $r(C,B) = 1$, $r(A,C) = 10$, $r(C,A) = 2$, discount factor is 0.5. Find the fixed point of the value function for the policy $\pi(A) = B$ (if we are in state A we choose the action to go to B) $\pi(B) = C$.$\pi(C) = A$. $v^{\pi}([A\ B\ C]) = ?$ *1 point*

- [6 6 6]
- [6.28 6.57 5.14]
- [6.34 6.88 5.92]
- [5.47 6.23 4.13]

No, the answer is incorrect.
Score: 0
Accepted Answers:
[6.28 6.57 5.14]

9) Which of the following is a contraction? *1 point*

- $T([x_1, x_2]) = [0.5x_1, 0.5x_2]$
- $T([x_1, x_2]) = [2x_1, 2x_2]$
- $T([x_1, x_2]) = x_1 + x_2$
- $T([x_1, x_2]) = [x_1, x_2, x_1 + x_2]$

No, the answer is incorrect.
Score: 0
Accepted Answers:
$T([x_1, x_2]) = [0.5x_1, 0.5x_2]$
$T([x_1, x_2]) = x_1 + x_2$

10) For an operator L, which of the following properties must be satisfied by x for it to be a fixed point for L? *1 point*

- $Lx = x$
- $L^2 x = x$
- $\forall \lambda > 0 \;\; Lx = \lambda x$
- None of the above

No, the answer is incorrect.
Score: 0
Accepted Answers:
$Lx = x$
$L^2 x = x$