Unit 12 - Week 10

Assignment 10

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

Due on 2020-04-08, 23:59 IST.

1) Which of the following is true about Markov and Semi Markov Options? *1 point*

- In a Markov Option the option's policy depends only on the current state
- In a Semi Markov Option the option's policy can depend only on the current state
- In a Semi Markov Option, the option's policy may depend on the history since the exe- cution of the option began
- A Semi-Markov Option is always a Markov Option but not vice versa

No, the answer is incorrect.
Score: 0

Accepted Answers:
In a Markov Option the option's policy depends only on the current state
In a Semi Markov Option the option's policy can depend only on the current state
In a Semi Markov Option, the option's policy may depend on the history since the exe- cution of the option began

2) What type of solution will you get if you try to find the optimal solution consistent with the hierarchical structure of the problem *1 point*

- Hierarchically optimal solution
- Recursively optimal solution
- Flat optimal solution

No, the answer is incorrect.
Score: 0

Accepted Answers:
Hierarchically optimal solution

3) Which if the following is a correct Bellman equation for an SMDP? *1 point*
note: $R(s, a, s') \implies$ reward is a function of only s, a and s'

- $V^*(s) = max_{a \in A(S)}[R(s, a, \tau, s') + \gamma^\tau P(s'|s, a)V^*(s')]$
- $V^*(s) = max_{a \in A(S)}[\Sigma_{s',a} P(s'|s, a, \tau)(R(s, a, \tau, s') + \gamma V^*(s'))]$
- $V^*(s) = max_{a \in A(S)}[\Sigma_{s',a} P(s', \tau|s, a)(R(s, a, \tau, s') + \gamma^\tau V^*(s'))]$
- $V^*(s) = max_{a \in A(S)}[\Sigma_{s',a} P(s', \tau|s, a)(R(s, a, s') + \gamma V^*(s'))]$

No, the answer is incorrect.
Score: 0

Accepted Answers:
$V^*(s) = max_{a \in A(S)}[\Sigma_{s',a} P(s', \tau|s, a)(R(s, a, \tau, s') + \gamma^\tau V^*(s'))]$

4) In Hierarchy of Abstract Machine the core MDP state changes only when we visit a choice state *1 point*

- True
- False

No, the answer is incorrect.
Score: 0

Accepted Answers:
False

5) Consider a SMDP in which the next state and the reward only depend on the previous state and action i.e $P(s', \tau|s, a) = P(s'|s, a)P(\tau|s, a)$, $R(s, a, \tau, s') = R(s, a, s')$. *1 point*
If we solve the above SMDP with conventional Q-learning we will end up with the same policy as solving it with SMDP Q-learning.

- yes, because now τ won't change anything and we end up with same states and action sequences
- no, because τ still depends on the action and discounting may have a effect on the final policies
- no, because the next state will still depend on the τ
- yes, because the bellman equation is same for both methods in this case

No, the answer is incorrect.
Score: 0

Accepted Answers:
no, because τ still depends on the action and discounting may have a effect on the final policies

6) In the above SMDP if τ is fixed for all states, actions will we always get the same policy for conventional Q-learning and SMDP Q learning then? For $\tau = 3, \tau = 2, \tau = 1$ *1 point*

- yes, yes, no
- no, no, no
- yes, yes, yes
- no, no, yes

No, the answer is incorrect.
Score: 0

Accepted Answers:
no, no, yes

7) Does flat optimal solutions gives the solution with highest expected returns *1 point*

- yes
- no

No, the answer is incorrect.
Score: 0

Accepted Answers:
yes

8) Suppose a deterministic markov option o_1 is being executed, and that the current state is s_1. The action taken in s_1 is a_1 which is in accordance to o_1. There exist 2 other deterministic markov options o_2, o_3. o_2 takes a_2 when in state s_1, and o_3 takes a_1 when in s_1. Which of the following values can be updated after the reward from taking action a_1 in state s_1 while executing option o_1 is obtained? *1 point*

- $Q(s_1, a_1)$
- $Q(s_1, o_1)$
- $Q(s_1, o_2)$
- $Q(s_1, o_3)$

No, the answer is incorrect.
Score: 0

Accepted Answers:
$Q(s_1, a_1)$
$Q(s_1, o_1)$
$Q(s_1, o_3)$

9) In HAM what will be the immediate rewards received between two choice states *1 point*

- Accumulation of immediate rewards of the core MDP obtained between these choice points
- The return of the next choice state
- The reward of only the next primitive action taken
- Immediate reward is always zero

No, the answer is incorrect.
Score: 0

Accepted Answers:
Accumulation of immediate rewards of the core MDP obtained between these choice points