Unit 10 - Week 8

Assignment 8

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

Due on 2020-03-25, 23:59 IST.

1) In which of the following cases would you expect to find that defining the loss of a function appoximator as $\sum_{s\in S}(\hat{V}(s) - V(s))^2$ leads to poor performance. Consider 'relevant' states to be those which are visited frequently when executing near optimal policies *1 point*

- Small state space, large percentage of relevant states
- Large state space, large percentage of relevant states.
- Large state space, small percentage of relevant states
- None of the above

No, the answer is incorrect.
Score: 0
Accepted Answers:
Large state space, small percentage of relevant states

2) **Assertion:** It is not possible to use look-up table based methods to solve continuous state space or continuous action space problems.(Assume discretization of continuous space is not allowed) *1 point*
Reason: A table which stores the Q value estimate for every state action pair can be built, however the look up time is too large for practical purposes.

- 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
- Both Assertion and Reason are false

No, the answer is incorrect.
Score: 0
Accepted Answers:
Assertion is true, Reason is false

3) Which of the following is the correct update equation for eligibility traces with linear function approximator, $V(s_t) = w^\top \phi(s_t)$? *1 point*

- $\vec{e}_t = \gamma\lambda\vec{e}_{t-1} + w$
- $\vec{e}_t = \gamma\lambda\vec{e}_{t-1} + \phi(s_t)$
- $\vec{e}_t = \gamma\lambda\vec{e}_{t-1} + \phi(s_{t-1})$
- None of the above

No, the answer is incorrect.
Score: 0
Accepted Answers:
$\vec{e}_t = \gamma\lambda\vec{e}_{t-1} + \phi(s_t)$

4) **Assertion:**To solve the given optimization problem for some states with linear function approximator, *1 point*

$$\pi_{t+1}(s) = \underset{a}{\text{argmax}}\ \hat{Q}^{\pi_t}(s,a)$$

we formulate a classification problem for discrete action space.
Reason: The given problem is equivalent to solving:

$$\pi_{t+1} = \underset{a}{\text{argmax}}\ \Phi\hat{\Theta}^{\pi_t}$$

which can be solved with categorical output, a and elements of $\Phi(s)$ as input

- 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
- Both Assertion and Reason are false

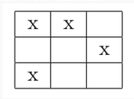
No, the answer is incorrect.
Score: 0
Accepted Answers:
Both Assertion and Reason are true, and Reason is correct explanation for Assertion

5) Which of the following is true about the LSTD and LSTDQ algorithm? *1 point*

- LSTD estimates matrices $A = \mathbb{E}[x_t(x_t - \gamma x_{t+1})^\top]$ and $b = \mathbb{E}[R_{t+1}x_t]$ and computes $\mathbf{w_{TD}} = A^{-1}b$
- LSTDQ approximates functions which are linear in feature vector
- Both LSTD and LSTDQ can reuse samples
- Both LSTD and LSTDQ are linear function approximation methods

No, the answer is incorrect.
Score: 0
Accepted Answers:
LSTD estimates matrices $A = \mathbb{E}[x_t(x_t - \gamma x_{t+1})^\top]$ and $b = \mathbb{E}[R_{t+1}x_t]$ and computes $\mathbf{w_{TD}} = A^{-1}b$
LSTDQ approximates functions which are linear in feature vector
Both LSTD and LSTDQ are linear function approximation methods

6) Consider the following small part of a gridworld. Say blocks with x in it are inaccessible, and empty blocks are accessible. And we want to do linear parameterization for states by denoting information of neighbouring states. Say the parameterization for state is as follows. '0' shows that the state is accessible, and '1' shows that it is inaccessible. For a state's vector, we define it as at vector of status' of its neigbours in the order (top, right, bottom, left). For the following sub-gridworld, what will be the state vector for the middle state? *1 point*



- (0,0,1,1)
- (1,0,0,1)
- (1,1,0,0)
- (0,1,1,0)

No, the answer is incorrect.
Score: 0
Accepted Answers:
(1,1,0,0)

7) Suppose we have a grid-world problem where the number of states is very large(say in millions). So, we define in some way clusters of neighbouring states and for learning, will assume that states within a cluster will have same value function. Assume that clusters are defined in such a way that all states belong to some cluster. Say true/false for the following statement. *1 point*
Statement: A single state can belong to more than one clusters.

- True
- False

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

8) Tile coding is a method of state aggregation for gridworld problems. Consider the following statements. *1 point*
i the number of indicators for each state is equal to number of tilings
ii Number of tilings for gridworld must always be 2
iii Tile coding is also a form of Coarse coding
Say which of the above statements are true

- iii only
- i, iii
- i only
- i, ii, iii

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

9) Which of the following are the correct values for $\tilde{A}$ in LSTDQ method. *1 point*
Note the samples are $D = \{s_i, a_i, s'_i, r_i\}$, when a policy ϕ is followed.

- $1/L\ \Sigma_{i=1}^{L}[\phi(s_i,a_i)(\phi(s_i,a_i) - \gamma\phi(s'_i,a_i))^T]$
- $1/L\ \Sigma_{i=1}^{L}[\phi(s_i)(\phi(s_i) - \gamma\phi(s'_i))^T]$
- $1/L\ \Sigma_{i=1}^{L}[\phi(s_i,a_i)(\phi(s_i,a_i) - \gamma\phi(s'_i))^T]$
- $1/L\ \Sigma_{i=1}^{L}[\phi(s_i,a_i)(\phi(s_i,a_i) - \gamma\phi(s'_i,\pi(s'_i)))^T]$

No, the answer is incorrect.
Score: 0
Accepted Answers:
$1/L\ \Sigma_{i=1}^{L}[\phi(s_i,a_i)(\phi(s_i,a_i) - \gamma\phi(s'_i,\pi(s'_i)))^T]$

10) In the LSTD method if number of samples $L \to \infty$ will $\Phi\theta^{\pi}$ always tend to the true returns of the underlying MDP, for the policy pi? *1 point*

- yes
- no

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