RL class - quiz on chapter 2

Total des points 3/7

Student last name *

Igbida

Student first name *

Rayanne

Question 1 * 1/1

Pick the answer(s) that correspond to Bellman's evaluation equation.

$$V^*(s) = \max_\pi \mathbb{E}[\sum_t \gamma^t R_t | \pi], \ \forall s \in S$$

☐ Option 1

$$Q(s,a) = \mathbb{E}_{r,s'}[r + \max'_a Q(s',a')], \quad \forall s,a \in S \times A$$

☐ Option 2

$$V(s) = \max_a[r(s,a) + \gamma \sum_{s'} p(s'|s,a)V(s')], \forall s \in S$$

☐ Option 3

$$Q(s,a) = r(s,a) + \gamma \int_{S \times A} Q(s',a')\pi(a'|s')p(s'|s,a)ds'da', \quad \forall s,a \in S \times A$$

☑ Option 4 ✓

Question 2 *0/1

Value iteration computes a sequence of value functions. Pick the answer(s) that correctly describe this sequence (notations are those used in class).

$$V_{n+1} = T^{\pi} V_n$$

- [ ] Option 1

$$Q_{n+1} = T^{*} Q_n$$

- [ ] Option 2

$$Q_{n+1} = T^{\pi_n} Q_n \text{ with } \pi_n \in \mathcal{G} Q_n$$

- [x] Option 3 ✓

$$V_{n+1} = T^{\pi_n} V_n \text{ with } \pi_n \in \mathcal{G} V_n$$

- [ ] Option 4

Bonne réponse

- [x] Option 2
- [x] Option 3

Question 3

*0/1

In MDPs with finite state-action spaces, value iteration (VI) can be written as a sequence of Bellman backups in each state or state-action pair. Pick the correct form(s) and statement(s) for these backups.

$Q(s,a) \leftarrow \sum_{s'} p(s'|s,a)[r(s,a,s')+ \gamma \max_{a'} Q(s',a')]$

- [ ] Option 1

$Q(s,a) \leftarrow r(s,a)+ \gamma \sum_{s'} p(s'|s,\pi(s))Q(s',\pi(s'))$

- [x] Option 2 ✗

$V(s) \leftarrow \sum_a \pi(a|s)[r(s,a) + \gamma \sum_{s'} V(s')]$

- [x] Option 3 ✗

When applied to V functions, one iteration of VI makes $|S|$ backups and each has complexity $|S||A|$.

- [ ] Option 4

When applied to Q functions, one iteration of VI makes $|S||A|$ backups and each has complexity $|S||A|$.

- [ ] Option 5

When applied to Q functions, one iteration of VI makes $|S||A|$ backups and each has complexity $|S|$.

- [ ] Option 6

Bonne réponse

- [x] Option 1

- [x] Option 4
- [x] Option 6

Question 4 *0/1

Modified policy iteration (MPI) computes a sequence of value functions and policies. Pick the true statement(s) about this sequence or about MPI.

$Q_{n+1} = (T^{\pi_n})^m Q_n$ and $\pi_n \in \mathcal{G}Q_n$

- [ ] Option 1

Policy iteration is a special case of MPI

- [x] Option 2

Value iteration is MPI when $m \to \infty$

- [ ] Option 3

$Q_{n+1} = (T^*)^m Q_n$ and $\pi_n \in \mathcal{G}Q_n$

- [ ] Option 4

Bonne réponse

- [x] Option 1
- [x] Option 2

Question 5 *

0/1

Pick the true statement(s) about greedy policies.

Any policy that is greedy with respect to Q^{π} is optimal.

- [x] Option 1 ✗

All policies in $\mathcal{G}Q^*$ are optimal.

- [ ] Option 2

There is always only a unique policy in $\mathcal{G}Q^*$.

- [ ] Option 3

$$Q^* = \mathcal{G}Q^*.$$

- [ ] Option 4

Bonne réponse

- [x] Option 2

Question 6 *

1/1

Pick the true statement(s) about policy iteration (PI).

PI alternates one application of T^π and one application of $\mathcal{G}$.

- [ ] Option 1

PI only stores value functions and does not require storing a policy.

- [ ] Option 2

$\mathcal{G}Q^{\pi_n}$ is always at least as good as π_n.

- [x] Option 3

PI fully solves the evaluation equation at each iteration.

- [x] Option 4

Question 7

*1/1

Asynchronous dynamic programming algorithms generalize VI, PI and MPI in discrete state and action MDPs. Pick the true statement(s).

MPI cannot be written as an asynchronous dynamic programming algorithm.

- [ ] Option 1

Gauss-Seidl VI always converges faster than VI.

- [ ] Option 2

Convergence of asynchronous DP algorithms requires that all states and actions are backed up an infinite number of times when $t \to \infty$.

- [x] Option 3 ✓

Asynchronous DP may enable finding π^* and Q^* faster than their synchronous counterparts.

- [x] Option 4 ✓