[Course](#) › [Unit 5 Reinforcement Learning (2 weeks)](#) › [Lecture 17. Reinforcement Learning 1](#) › 3. RL Terminology

Audit Access Expires May 11, 2020

You lose all access to this course, including your progress, on May 11, 2020.

3. RL Terminology

RL Terminology

Video

[Download video file](#)

Transcripts

[**Download SubRip (.srt) file**](#)

[**Download Text (.txt) file**](#)

A **Markov decision process (MDP)** is defined by

- a set of states $s \in S$;
- a set of actions $a \in A$;
- Action dependent transition probabilities $T(s, a, s') = P(s'|s, a)$, so that for each state s and action a, $\sum_{s' \in S} T(s, a, s') = 1$.
- Reward functions $R(s, a, s')$, representing the reward for starting in state s, taking action a and ending up in state s' after one step. (The reward function may also depend only on s, or only s and a.)

MDPs satisfy the **Markov property** in that the transition probabilities and rewards depend only on the current state and action, and remain unchanged regardless of the history (i.e. past states and actions) that leads to the current state.

Note: If you have taken *6.431x Probability–The Science of Uncertainty and Data*, you may review the Markov Property and Markov Chains introduced in Unit 10 Markov Processes. Markov decision processes are extensions of Markov Chains by the set of actions (and the action-dependence of the transition probabilities), and the reward functions.

Markov Property

1/1 point (graded)

Let $X_i, i = 1, 2, \ldots$ be a discrete Markov chain with states $\{s_j, j \in \mathbb{N}\}$. Which of the following statements are correct?

☑ For $n \geq 3$,
$P\left[X_n = x_n \mid X_{n-1} = x_{n-1}, X_1 = x_1\right] = P\left[X_n = x_n \mid X_{n-1} = x_{n-1}\right].$

☑ For $n \geq 3$ and $n - j > 1$,
$P\left[X_n = x_n \mid X_{n-j} = x_{n-j}, X_1 = x_1\right] = P\left[X_n = x_n \mid X_{n-j} = x_{n-j}\right].$

□ For $n \geq 2$,
$P\left[X_n = x_n \mid X_{n+1} = x_{n+1}, X_1 = x_1\right] = P\left[X_n = x_n \mid X_1 = x_1\right].$

✓

Solution:

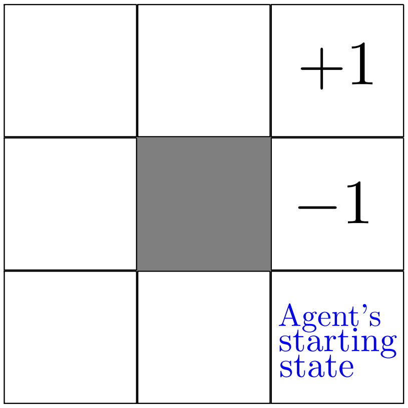
The first two statements are a direct consequence of the Markov property. The third choice is not correct (see if you can come up with a quick counter-example).

Submit

You have used 2 of 2 attempts

ⓘ Answers are displayed within the problem

Recall the MDP from the lecture.



An AI agent navigates in the 3x3 grid depicted above, where the middle square is not accessible by the agent (and hence is greyed out).

The MDP is defined as follows:

- Every state s is defined by the current position of the agent in the grid (and is independent of its previous actions and positions).
- The actions a are the 4 directions “up", “down",“left", “right".
- The transition probabilities from state s via action a to state s' is given by $T(s,a,s') = P(s'|s,a)$.
- The agent receives a reward of $+1$ for arriving at the top right cell, and a reward of -1 for arriving in the cell immediately below it. It does not receive any non-zero reward at the other cells as illustrated in the following figure.

Markovian Setting

0/1 point (graded)

Let s be any given state in this MDP. The agent takes actions $a_1, a_2 \ldots a_n$ starting from state s_0 and as a result visits states $s_1, s_2 \ldots s_n = s$ in that order.

Given that $s_n = s,$ that is, the agent ends up at the current state s after n steps, what do the rewards after the n^{th} step depend on? (Choose all that apply.)

☑ Rewards collected after the n^{th} step do not depend on the previous states $s_1, s_2 \ldots s_{n-1}$ ✓

□ Rewards collected after the n^{th} step can depend on the previous states $s_1, s_2 \ldots s_{n-1}$

□ Rewards collected after the n^{th} step can depend on the current state s ✓

☑ Rewards collected after the n^{th} step do not depend on the previous actions $a_1, a_2 \ldots a_n$ ✓

✗

Solution:

Note that under a Markovian setting, the rewards and the state transition probabilities given the current state would be independent of the previous states and actions. However, they would depend on the current state and the current action (s, a_{n+1} in our example).

Submit You have used 3 of 3 attempts

ℹ Answers are displayed within the problem

Number of States

1/1 point (graded)

Enter the total number of unique states in the MDP described above and depicted by the 3×3 grid above. (Enter -1 if the number of states is not finite.)

8 ✓ **Answer:** 8

Solution:

Each state corresponds to a unique position that the agent could be at. Since, the agent isn't allowed to be at the center of the grid, there are a total of 8 possible positions and hence the cardinality of the state space for this example is 8.

Submit You have used 1 of 3 attempts

ℹ Answers are displayed within the problem

Transition Probabilities

0/1 point (graded)

Refer to the MDP described and depicted in the 3×3 grid on the top of this page.

Assume that the transition probabilities for all the states are given as a table M, whose (i, j, k)-th entry is $M[i][j][k] = T(s_i, a_j, s_k) = P(s_k | s_i, a_j)$, which represents the transition probability of ending up at state s_k when action a_j is taken from the state s_i. **Note:** Note that here, s_i and s_k can be any pair of states, not necessarily reachable by an action in one step.

Enter the number of entries in the table M:

28

✖ **Answer:** 256

Solution:

Note that the transition probability table has a probability value $P(s'|s,a)$ associated with each of the tuples (s, a, s') where $P(s'|s,a)$ is the probability of reaching state s' if the agent chooses action a at state s.
Since there are 8 states and 4 actions, the size of this table would be
$8 * 8 * 4 = 256$.
Also note that for any given state, action pair (s, a), the following must hold

$$\sum_{s'} P(s'|s,a) = 1$$

Submit

You have used 3 of 3 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 5 Reinforcement Learning (2 weeks) :Lecture 17. Reinforcement Learning 1 / 3. RL Terminology

Add a Post

Show all posts

by recent activity

- 💬 https://www.youtube.com/watch?v=cPUfh2rRJik 4
  https://www.youtube.com/watch?v=cPUfh2rRJik might be helpful to some
  📌 Pinned

- 💬 missed completing the submission. 3
  missed completing the submission today early morning though i have saved my answers whi...

- ☑ More Transition Probabilities Clarification 3
  Do we consider the +1 and/or the -1 states to be halting states? And if so, would halting state...

- ☑ Number of entries in table M 15
  How to represent in the table, the case when multiple actions are needed to transition from ...

- ? Did you use Bayes Theorem to show that third answer is incorrect for the Markov Property question? 1
  I solved a small expression to show that third part doesn't satisfy the Markov property

- 💬 Markovian setting: not only transition and rewards but also behaviour 1
  "MDPs satisfy the Markov property in that the transition probabilities and rewards depend o...
  👤 Community TA

- ? Markovian Setting (do not depend on the previous actions / states) 10
  Generally I got the idea to which the question is pointing, however one thing is not entirely cl...

- ? Markovian Setting 1

- ☑ How the reward is modelled ? 2
  👤 Community TA

- ? Transition Probabilities: asking for clarifications 5

- ☑ [Staff] Number of entries in M 10
  Do all the cells of the tables would have non zero entries? I think 28 out of N cells (N is the nu...

- ☑ Why is s' not determined by (s,a)? 5
  I must be missing something very obvious here but why isn't s' fully determined by the curre...