Congratulations! You passed!

Grade received 100% | Latest Submission Grade 100% | To pass 80% or higher

Go to next item

1. TD(0) is a solution method for: **1 / 1 point**

- ○ Control
- ◉ Prediction

Correct
Correct! TD(0) is used to estimate the value function for a given policy. In other words, it is a solution method for the prediction problem.

2. Which of the following methods use bootstrapping? (Select all that apply) **1 / 1 point**

- ☑ Dynamic Programming

Correct
Correct! DP algorithms are obtained by turning Bellman equations into update rules for improving approximations of the desired value functions. These methods update estimates of the values of states based on estimates of the values of successor states. That is, they update estimates on the basis of other estimates.

- ☐ Monte Carlo
- ☑ TD(0)

Correct
Correct! Temporal Difference methods update "a guess from a guess". They estimate the value of the current state using the immediate reward and the estimate of the value in the next state. They bootstrap- off their own estimates.

3. Which of the following is the correct characterization of Dynamic Programming (DP) and Temporal Difference (TD) methods? **1 / 1 point**

- ○ Both TD and DP methods use *expected* updates.
- ○ Both TD and DP methods use *sample* updates.
- ○ TD methods use *expected* updates, DP methods use *sample* updates.
- ◉ TD methods uses *sample* updates, DP methods use *expected* updates.

Correct
Correct! TD methods use samples to update value estimates. On the other hand, Dynamic Programming methods use a model to perform expected updates.

4. Match the algorithm name to its correct update (**select all that apply**) **1 / 1 point**

- ☑ Monte Carlo: $V(S_t) \leftarrow V(S_t) + \alpha[G_t - V(S_t)]$

Correct
Correct! Monte-Carlo methods update value estimates toward empirically observed returns.

- ☐ Monte Carlo: $V(S_t) \leftarrow V(S_t) + \alpha[R_{t+1} + \gamma V(S_{t+1}) - V(S_t)]$
- ☐ TD(0): $V(S_t) \leftarrow V(S_t) + \alpha[G_t - V(S_t)]$
- ☑ TD(0): $V(S_t) \leftarrow V(S_t) + \alpha[R_{t+1} + \gamma V(S_{t+1}) - V(S_t)]$

Correct
Correct! TD(0) updates value estimates toward the TD(0)-target of the sum of the observed reward and discounted next state value.

5. Which of the following well-describe Temporal Difference (TD) and Monte-Carlo (MC) methods? **1 / 1 point**

- ☑ TD methods can be used in *continuing* tasks.

Correct
Correct! The returns in continuing tasks are sums of rewards infinitely into the future. But, TD does not have to wait to get samples of these returns. The targets can be obtained immediately, using bootstrapping.

- ☐ MC methods can be used in *continuing* tasks.
- ☑ TD methods can be used in *episodic* tasks.