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

- ◯ Control
- ◯ Prediction

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

- ☐ Dynamic Programming
- ☐ Monte Carlo
- ☐ TD(0)

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

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

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

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

- [ ] TD methods can be used in *continuing* tasks.
- [ ] MC methods can be used in *continuing* tasks.
- [ ] TD methods can be used in *episodic* tasks.
- [ ] MC methods can be used in *episodic* tasks.

6. In an episodic setting, we might have different updates depending on whether the next state is terminal or non-terminal. Which of the following TD error calculations are correct? **1 point**

- [ ] S_{t+1} is non-terminal: $\delta_t = R_{t+1} + \gamma V(S_{t+1}) - V(S_t)$
- [ ] S_{t+1} is non-terminal: $\delta_t = R_{t+1} - V(S_t)$
- [ ] S_{t+1} is terminal: $\delta_t = R_{t+1} + \gamma V(S_{t+1}) - V(S_t)$ with $V(S_{t+1}) = 0$
- [ ] S_{t+1} is terminal: $\delta_t = R_{t+1} - V(S_t)$

7. Suppose we have current estimates for the value of two states: V(A) = 1.0, V(B) = 1.0 in an episodic setting. We observe the following trajectory: A, 0, B, 1, B, 0, T where T is a terminal state. Apply TD(0) with step-size, $\alpha = 1$, and discount factor, $\gamma = 0.5$. What are the value estimates for state A and state B at the end of the episode? **1 point**

- ◯ (1.0, 1.0)
- ◯ (0.5, 0)
- ◯ (0, 1.5)
- ◯ (1, 0)
- ◯ (0, 0)

8. Which of the following pairs is the correct characterization of the targets used in TD(0) and Monte Carlo? **1 point**

- ◯ TD(0): High Variance Target, Monte Carlo: High Variance Target
- ◯ TD(0): High Variance Target, Monte Carlo: Low Variance Target

- TD(0): Low Variance Target, Monte Carlo: High Variance Target
- TD(0): Low Variance Target, Monte Carlo: Low Variance Target

9. Suppose you observe the following episodes of the form (State, Reward, ...) from a Markov Decision Process with states A and B: **1 point**

Episodes
A, 0, B, 0
B, 1
B, 1
B, 1
B, 0
B, 0
B, 1
B, 0

What would batch Monte Carlo methods give for the estimates V(A) and V(B)? What would batch TD(0) give for the estimates V(A) and V(B)? Use a discount factor, γ, of 1.

For Batch MC: compute the average returns observed from each state. For Batch TD: You can start with state B. What is its expected return? Then figure out V(A) using the temporal difference equation: $V(S_t) = E[R_{t+1} + \gamma V(S_{t+1})]$.

Answers are provided in the following format:

- $V^{\text{batch-MC}}(A)$ is the value for state A under Monte Carlo learning
- $V^{\text{batch-MC}}(B)$ is the value of state B under Monte Carlo learning
- $V^{\text{batch-TD}}(A)$ is the value of state A under TD learning
- $V^{\text{batch-TD}}(B)$ is the value of state B under TD learning

Hint: review example 6.3 in Sutton and Barto; this question is the same, just with different numbers.

○ $V^{\text{batch-MC}}(A) = 0$

$V^{\text{batch-MC}}(B) = 0.5$

$V^{\text{batch-TD}}(A) = 0.5$

$V^{\text{batch-TD}}(B) = 0.5$

- ○ $V^{\text{batch-MC}}(A) = 0$

  $V^{\text{batch-MC}}(B) = 0.5$

  $V^{\text{batch-TD}}(A) = 0$

  $V^{\text{batch-TD}}(B) = 0.5$
- ○ $V^{\text{batch-MC}}(A) = 0$

  $V^{\text{batch-MC}}(B) = 0.5$

  $V^{\text{batch-TD}}(A) = 0$

  $V^{\text{batch-TD}}(B) = 0$
- ○ $V^{\text{batch-MC}}(A) = 0$

  $V^{\text{batch-MC}}(B) = 0.5$

  $V^{\text{batch-TD}}(A) = 1.5$

  $V^{\text{batch-TD}}(B) = 0.5$
- ○ $V^{\text{batch-MC}}(A) = 0.5$

  $V^{\text{batch-MC}}(B) = 0.5$

  $V^{\text{batch-TD}}(A) = 0.5$

  $V^{\text{batch-TD}}(B) = 0.5$

10. True or False: "Both TD(0) and Monte-Carlo (MC) methods converge to the true value function asymptotically, given that the environment is Markovian." **1 point**

- ○ True
- ○ False

11. Which of the following pairs is the correct characterization of the TD(0) and Monte-Carlo (MC) methods? **1 point**

- ○ Both TD(0) and MC are offline methods.
- ○ Both TD(0) and MC are online methods.

- TD(0) is an online method while MC is an offline method.
- MC is an online method while TD(0) is an offline method.