Congratulations! You passed!

Grade received 81.81%

Latest Submission Grade 81.82%

To pass 80% or higher

Go to next item

1. A function which maps ___ to ___ is a value function. [Select all that apply]

1 / 1 point

□ Values to actions.

□ Values to states.

☑ State-action pairs to expected returns.

Correct
Correct! A function that takes a state-action pair and outputs an expected return is a value function.

☑ States to expected returns.

Correct
Correct! A function that takes a state and outputs an expected return is a value function.

2. Consider the continuing Markov decision process shown below. The only decision to be made is in the top state, where two actions are available, left and right. The numbers show the rewards that are received deterministically after each action. There are exactly two deterministic policies, π_{left} and π_{right}. Indicate the optimal policies if $\gamma = 0$? If $\gamma = 0.9$? If $\gamma = 0.5$? [Select all that apply]

1 / 1 point

☑ For $\gamma = 0, \pi_{\text{left}}$

Correct
Correct! Since both policies return to the top state every two time steps, to determine the optimal policy, it suffices to consider the reward accumulated over the first two time steps. For the policy left, this is equal to 1; for the policy right, this is equal to 0.

☑ For $\gamma = 0.9, \pi_{\text{right}}$

Correct
Correct! Since both policies return to the top state every two time steps, to determine the optimal policy, it suffices to consider the reward accumulated over the first two time steps. For the policy left, this is equal to 1; for the policy right, this is equal to 1.8.

☑ For $\gamma = 0.5, \pi_{\text{left}}$

Correct
Correct! Since both policies return to the start state every two time steps, to determine the optimal policy, it suffices to consider the reward accumulated over the first two time steps. For the policy left, this is equal to 1; for the policy right, this is equal to 1.

□ For $\gamma = 0.9, \pi_{\text{left}}$

☑ For $\gamma = 0.5, \pi_{\text{right}}$

Correct
Correct! Since both policies return to the start state every two time steps, to determine the optimal policy, it suffices to consider the reward accumulated over the first two time steps. For the policy left, this is equal to 1; for the policy right, this is equal to 1.

□ For $\gamma = 0, \pi_{\text{right}}$

3. Every finite Markov decision process has __. [Select all that apply]

1 / 1 point

□ A unique optimal policy

☑ A deterministic optimal policy

Correct
Correct! Let's say there is a policy π_1 which does well in some states, while policy π_2 does well in others. We could combine these policies into a third policy π_3, which always chooses actions according to whichever of policy π_1 and π_2 has the highest value in the current state. π_3 will necessarily have a value greater than or equal to both π_1 and π_2 in every state! So we will never have a situation where doing well in one state requires sacrificing value in another. Because of this, there always exists some policy which is best in every state. This is of course only an informal argument, but there is in fact a rigorous proof showing that there must always exist at least one optimal deterministic policy.

☑ A unique optimal value function

Correct
Correct! The Bellman optimality equation is actually a system of equations, one for each state, so if there are N states, then there are N equations in N unknowns. If the dynamics of the environment are known, then in principle one can solve this system of equations for the optimal value function using any one of a