Congratulations! You passed!

Grade received 90.90%
Latest Submission Grade 90.91%
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**

- [x] States to expected returns.

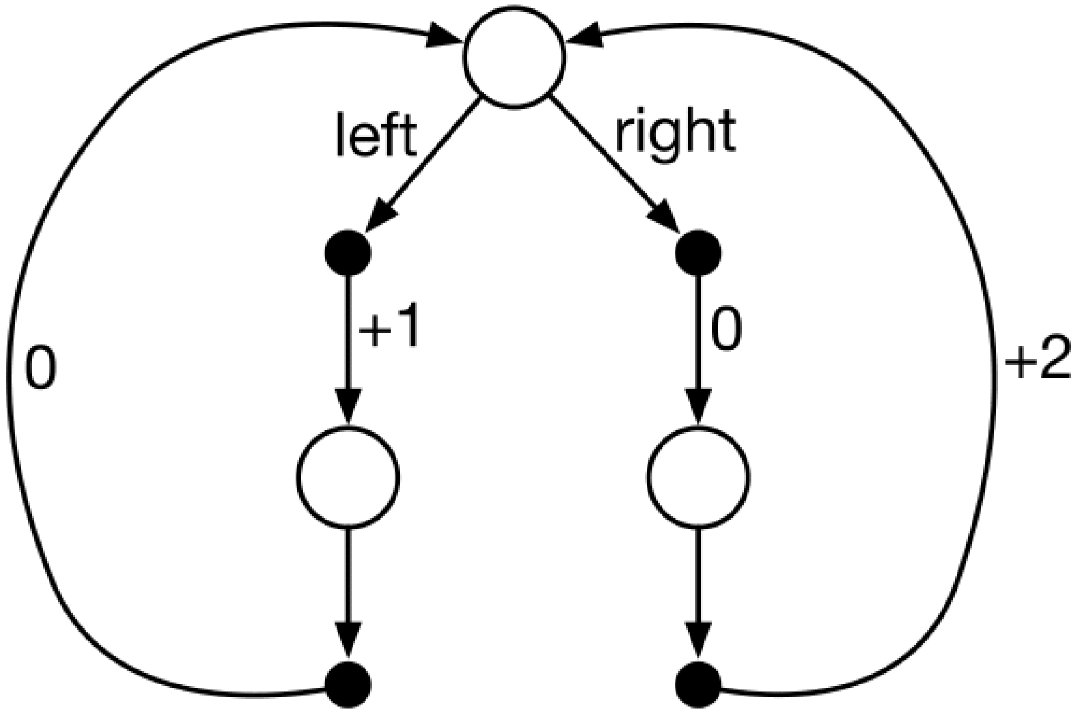
  Correct
  Correct! A function that takes a state and outputs an expected return is a value function.

- [ ] Values to states.
- [x] 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.

- [ ] Values to actions.

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 $\gamma = 0.9$, π_{left}
- [x] For $\gamma = 0.5$, π_{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 $\gamma = 0$, π_{right}
- [x] For $\gamma = 0$, π_{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.

- [x] For $\gamma = 0.5$, π_{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.

- [x] For $\gamma = 0.9$, π_{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.

3. Every finite Markov decision process has __. [Select all that apply] **0 / 1 point**

- [ ] A unique optimal value function
- [x] 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.

- [x] A unique optimal policy

This should not be selected
Incorrect. Take another look at the lesson: Optimal Policies.

- [ ] A stochastic optimal policy

4. The ___ of the reward for each state-action pair, the dynamics function p, and the policy π is _____ to characterize the value function v_π. (Remember that the value of a policy π at state s is $v_\pi(s) = \sum_a \pi(a|s) \sum_{s',r} p(s',r|s,a)[r + \gamma v_\pi(s')]$.)

1 / 1 point

- ○ Distribution; necessary
- ◉ Mean; sufficient

✓ **Correct**
Correct! If we have the expected reward for each state-action pair, we can compute the expected return under any policy.

5. The Bellman equation for a given a policy π: [Select all that apply]

1 / 1 point

- [x] Expresses state values $v(s)$ in terms of state values of successor states.

✓ **Correct**
Correct!

- [ ] Expresses the improved policy in terms of the existing policy.
- [ ] Holds only when the policy is greedy with respect to the value function.

6. An optimal policy:

1 / 1 point

- ◉ Is not guaranteed to be unique, even in finite Markov decision processes.
- ○ Is unique in every Markov decision process.
- ○ Is unique in every finite Markov decision process.

✓ **Correct**
Correct! For example, imagine a Markov decision process with one state and two actions. If both actions receive the same reward, then any policy is an optimal policy.

7. The Bellman optimality equation for v_*: [Select all that apply]

1 / 1 point

- [x] Expresses state values $v_*(s)$ in terms of state values of successor states.

Correct
Correct!

- [x] Holds for the optimal state value function.

Correct
Correct!

- [ ] Expresses the improved policy in terms of the existing policy.
- [ ] Holds when the policy is greedy with respect to the value function.
- [ ] Holds for v_π, the value function of an arbitrary policy π.

8. Give an equation for v_π in terms of q_π and π. **1 / 1 point**

- ( ) $v_\pi(s) = \sum_a \gamma\pi(a|s)q_\pi(s,a)$
- (•) $v_\pi(s) = \sum_a \pi(a|s)q_\pi(s,a)$
- ( ) $v_\pi(s) = \max_a \pi(a|s)q_\pi(s,a)$
- ( ) $v_\pi(s) = \max_a \gamma\pi(a|s)q_\pi(s,a)$

Correct
Correct!

9. Give an equation for q_π in terms of v_π and the four-argument p. **1 / 1 point**

- (•) $q_\pi(s,a) = \sum_{s'} \sum_r p(s',r|s,a)[r + \gamma v_\pi(s')]$
- ( ) $q_\pi(s,a) = \max_{s',r} p(s',r|s,a)[r + \gamma v_\pi(s')]$
- ( ) $q_\pi(s,a) = \sum_{s'} \sum_r p(s',r|s,a)[r + v_\pi(s')]$
- ( ) $q_\pi(s,a) = \sum_{s'} \sum_r p(s',r|s,a)\gamma[r + v_\pi(s')]$
- ( ) $q_\pi(s,a) = \max_{s',r} p(s',r|s,a)[r + v_\pi(s')]$
- ( ) $q_\pi(s,a) = \max_{s',r} p(s',r|s,a)\gamma[r + v_\pi(s')]$

Correct
Correct!

10. Let $r(s, a)$ be the expected reward for taking action a in state s, as defined in equation 3.5 of the textbook. Which of the following are valid ways to re-express the Bellman equations, using this expected reward function? [**Select all that apply**]

1 / 1 point

- [x] $q_\pi(s, a) = r(s, a) + \gamma \sum_{s'} \sum_{a'} p(s'|s, a)\pi(a'|s')q_\pi(s', a')$

  Correct

  Correct!

- [x] $q_*(s, a) = r(s, a) + \gamma \sum_{s'} p(s'|s, a) \max_{a'} q_*(s', a')$

  Correct

  Correct!

- [x] $v_\pi(s) = \sum_a \pi(a|s)[r(s, a) + \gamma \sum_{s'} p(s'|s, a)v_\pi(s')]$

  Correct

  Correct!

- [x] $v_*(s) = \max_a[r(s, a) + \gamma \sum_{s'} p(s'|s, a)v_*(s')]$

  Correct

  Correct!

11. Consider an episodic MDP with one state and two actions (left and right). The left action has stochastic reward 1 with probability p and 3 with probability $1 - p$. The right action has stochastic reward 0 with probability q and 10 with probability $1 - q$. What relationship between p and q makes the actions equally optimal?

1 / 1 point

- ( ) $7 + 2p = -10q$
- ( ) $13 + 3p = -10q$
- ( ) $13 + 3p = 10q$
- (•) $7 + 2p = 10q$
- ( ) $13 + 2p = 10q$
- ( ) $7 + 3p = -10q$
- ( ) $13 + 2p = -10q$

- $7 + 3p = 10q$

Correct
Correct!