

Congratulations! You passed!

TO PASS 80% or higher

GRADE

100%

Keep Learning

Dynamic Programming

TOTAL POINTS 14

1. The value of any state under an optimal policy is ___ the value of that state under a non-optimal policy. [Select all that apply]

1 / 1 point

- Strictly greater than
 Greater than or equal to
 Strictly less than
 Less than or equal to

Correct
 Correct! This follows from the policy improvement theorem.

2. If a policy is greedy with respect to the value function for the equiprobable random policy, then it is guaranteed to be an optimal policy.

1 / 1 point

- True
 False

Correct! Only policies greedy with respect to the optimal value function are guaranteed to be optimal.

3. If a policy π is greedy with respect to its own value function v_π , then it is an optimal policy.

1 / 1 point

- True
 False

Correct
 Correct! If a policy is greedy with respect to its own value function, it follows from the policy improvement theorem and the Bellman optimality equation that it must be an optimal policy.

4. Let v_π be the state-value function for the policy π . Let π' be greedy with respect to v_π . Then $v_{\pi'} \geq v_\pi$.

1 / 1 point

- False
 True

Correct
 Correct! This is a consequence of the policy improvement theorem.

5. Let v_π be the state-value function for the policy π . Let $v_{\pi'}$ be the state-value function for the policy π' . Assume $v_\pi = v_{\pi'}$. Then this means that $\pi = \pi'$.

1 / 1 point

- False
 True

Correct! For example, two policies might share the same value function, but differ due to random tie breaking.

6. What is the relationship between value iteration and policy iteration? [Select all that apply]

1 / 1 point

- Value iteration and policy iteration are both special cases of generalized policy iteration.

Correct
 Correct!

- Policy iteration is a special case of value iteration.
 Value iteration is a special case of policy iteration.

7. The word synchronous means "at the same time". The word asynchronous means "not at the same time". A dynamic programming algorithm is: [Select all that apply]

1 / 1 point

- Asynchronous, if it updates some states more than others.

Correct
 Correct! Only algorithms that update every state exactly once at each iteration are synchronous.

- Asynchronous, if it does not update all states at each iteration.

Correct
 Correct! Only algorithms that update every state exactly once at each iteration are synchronous.

8. All Generalized Policy Iteration algorithms are synchronous.

1 / 1 point

- True
 False

Correct
 Correct! A Generalized Policy Iteration algorithm can update states in a non-systematic fashion.

9. Policy iteration and value iteration, as described in chapter four, are synchronous.

1 / 1 point

- True
 False

Correct
 Correct! As described in lecture, policy iteration and value iteration update all states systematic sweeps.

10. Which of the following is true?

1 / 1 point

- Asynchronous methods generally scale to large state spaces better than synchronous methods.

- Synchronous methods generally scale to large state spaces better than asynchronous methods.

Correct

may be intractable.

11. Why are dynamic programming algorithms considered planning methods? [Select all that apply]

1 / 1 point

- They use a model to improve the policy.

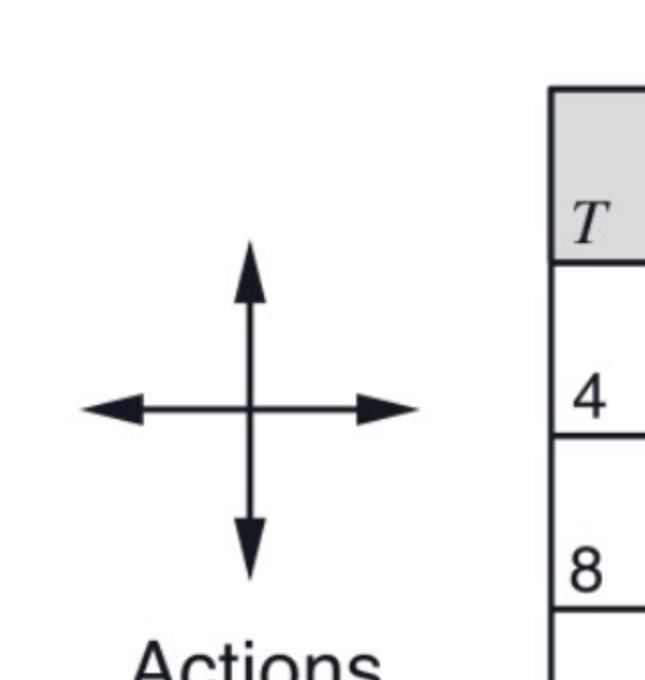
Correct
 Correct! This is the definition of a planning method.

- They compute optimal value functions.

- They learn from trial and error interaction.

12. Consider the undiscounted, episodic MDP below. There are four actions possible in each state, $A = \{\text{up}, \text{down}, \text{right}, \text{left}\}$, which deterministically cause the corresponding state transitions, except that actions that would take the agent off the grid in fact leave the state unchanged. The right half of the figure shows the value of each state under the equiprobable random policy. If π is the equiprobable random policy, what is $q(11, \text{down})$?

1 / 1 point



$R = -1$
 on all transitions

T	-14	-20	-22
-14	-18	-20	-20
-20	-20	-18	-14
-22	-20	-14	T

- $q(11, \text{down}) = -15$

- $q(11, \text{down}) = 0$

Correct

Correct! Moving down incurs a reward of -1 before reaching the terminal state, after which the episode is over.

13. Consider the undiscounted, episodic MDP below. There are four actions possible in each state, $A = \{\text{up}, \text{down}, \text{right}, \text{left}\}$, which deterministically cause the corresponding state transitions, except that actions that would take the agent off the grid in fact leave the state unchanged. The right half of the figure shows the value of each state under the equiprobable random policy. If π is the equiprobable random policy, what is $q(7, \text{down})$?

1 / 1 point

$R = -1$
 on all transitions

T	-14	-20	-22
-14	-18	-20	-20
-20	-20	-18	-14
-22	-20	-14	T

- $q(7, \text{down}) = -20$

- $q(7, \text{down}) = -15$

- $q(7, \text{down}) = -21$

- $q(7, \text{down}) = -24$

Correct

Correct! Moving down incurs a reward of -1 before reaching state 11, from which the expected future return is -14.

14. Consider the undiscounted, episodic MDP below. There are four actions possible in each state, $A = \{\text{up}, \text{down}, \text{right}, \text{left}\}$, which deterministically cause the corresponding state transitions, except that actions that would take the agent off the grid in fact leave the state unchanged. The right half of the figure shows the value of each state under the equiprobable random policy. If π is the equiprobable random policy, what is $v(15)$? Hint: Recall the Bellman equation $v(s) = \sum_a \pi(a|s) \sum_{s', r} p(s', r|s, a)[r + \gamma v(s')]$.

1 / 1 point

$R = -1$
 on all transitions

T	-14	-20	-22
-14	-18	-20	-20
-20	-20	-18	-14
-22	-20	-14	T

- $v(15) = -21$

- $v(15) = -22$

- $v(15) = -24$

- $v(15) = -25$

Correct

Correct! We get this by solving for the unknown variable $v(15)$. Let's call this unknown x .

We solve for x in the equation $x = 1/4(-21) + 3/4(-1 + x)$. The first term corresponds to