Congratulations! You passed!

Grade received 100% To pass 80% or higher

Go to next item

Dynamic Programming

Total points 10

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
- [x] Greater than or equal to

Correct
Correct! This follows from the policy improvement theorem.

- [ ] Strictly less than
- [ ] Less than or equal to

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

- (●) False
- ( ) True

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

3. 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.

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

1 / 1 point

- [ ] Policy iteration is a special case of value iteration.
- [x] Value iteration and policy iteration are both special cases of generalized policy iteration.

Correct
Correct!

- [ ] Value iteration is a special case of policy iteration.

5. 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

- [x] Synchronous, if it systematically sweeps the entire state space at each iteration.

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

- [x] 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.

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

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

6. 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.

7. 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.

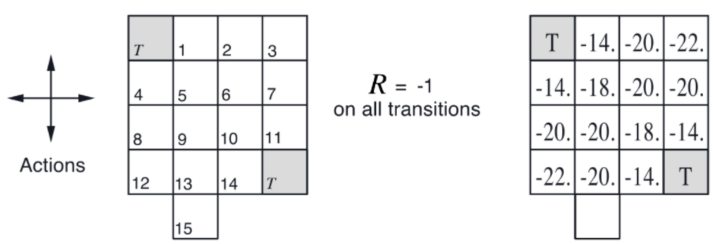
Correct
Correct! Asynchronous methods can focus updates on more relevant states, and update less relevant states less often. If the state space is very large, asynchronous methods may still be able to achieve good performance whereas even just one synchronous sweep of the state space may be intractable.

8. Why are dynamic programming algorithms considered planning methods? [Select all that apply] **1 / 1 point**

- [ ] They learn from trial and error interaction.
- [ ] They compute optimal value functions.
- [x] They use a model to improve the policy.

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

9. Consider the undiscounted, episodic MDP below. There are four actions possible in each state, A = {up, down, right, 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**

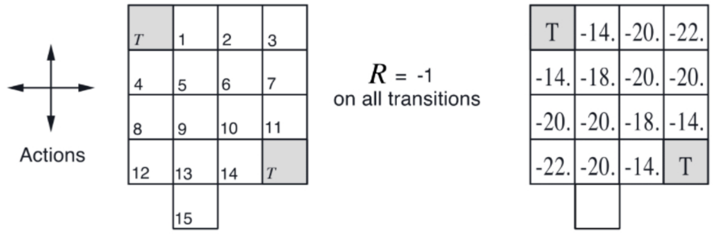


- ◉ $q(11, \text{down}) = -1$
- ○ $q(11, \text{down}) = -15$
- ○ $q(11, \text{down}) = -14$
- ○ $q(11, \text{down}) = 0$

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

10. Consider the undiscounted, episodic MDP below. There are four actions possible in each state, A = {up, down, right, 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



- ○ $v(15) = -23$
- ○ $v(15) = -22$
- ◉ $v(15) = -24$
- ○ $v(15) = -25$
- ○ $v(15) = -21$

Correct

Correct! We can 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 transitioning to state 13. The second term corresponds to taking one of the other three actions, incurring a reward of -1 and staying in state x.