Your grade: 90%

Your latest: 90% • Your highest: 90% • To pass you need at least 80%. We keep your highest score.

Next item

Dynamic Programming

coach

Ready to review what you've learned before starting the assignment? I'm here to help.

Help me practice Let's chat

Assignment details

Submitted	Attempts
Jan 12, 11:44 PM IST	Unlimited

Retry

Your grade

To pass you need at least 80%. We keep your highest score.

90%

View submission See feedback

Like Dislike Report an issue

---

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

Correct
Correct! This follows from the policy improvement theorem.

2. 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.

3. 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

- ( ) True
- (•) False

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

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

Correct
Correct!

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

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] 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.

6. 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.

7. Which of the following is true? — 1 / 1 point

- ( ) Synchronous methods generally scale to large state spaces better than asynchronous methods.
- (•) Asynchronous methods generally scale to large state spaces better than synchronous methods.

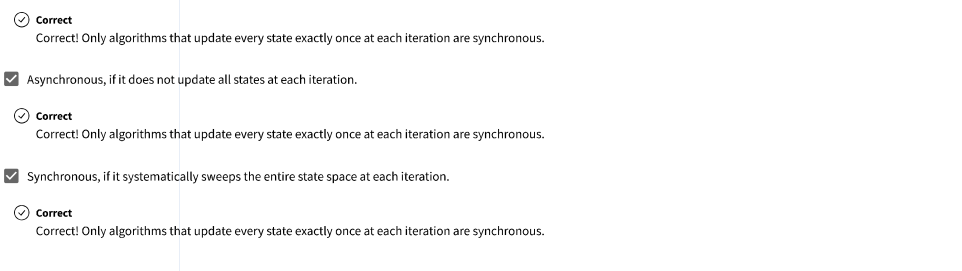
Correct
Correct! Asynchronous methods can focus updates on more relevant states, and update less relevant states less often. If the state space is very large, asynchronous methods may still be able to achieve good performance whereas even just one synchronous sweep of the state space may be intractable.

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

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

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

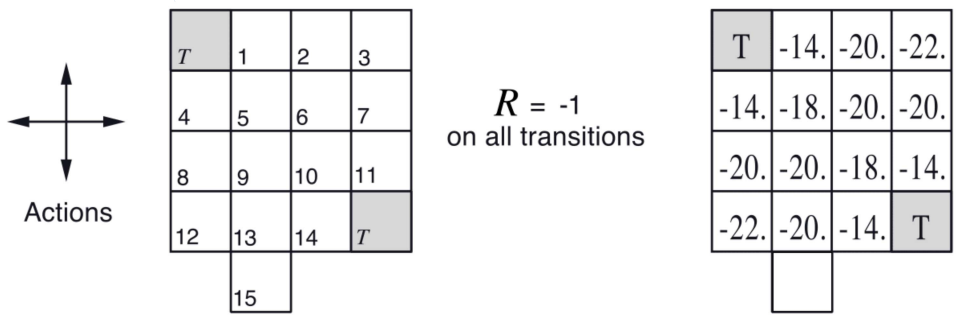
9. Consider the undiscounted, episodic MDP below. There are four actions possible in each state, A = {up, down, right, left}, which deterministically cause the corresponding state transitions, except that actions that would take the agent off the grid in fact leave the state unchanged. The right half of the figure shows the value of each state under the equiprobable random policy. If π is the equiprobable random policy, what is $q(7, \text{down})$? — 0 / 1 point



- ( ) $q(7, \text{down}) = -14$
- ( ) $q(7, \text{down}) = -20$
- (•) $q(7, \text{down}) = -21$
- ( ) $q(7, \text{down}) = -15$

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

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



- ( ) $v(15) = -21$
- ( ) $v(15) = -23$
- (•) $v(15) = -24$
- ( ) $v(15) = -25$