Dynamic Programming

Practice Quiz • 45 min • 10 total points

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 point**

- [ ] Strictly greater than
- [ ] Greater than or equal to
- [ ] 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 point**

- ○ False
- ○ True

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 point**

- ○ True
- ○ False

4. What is the relationship between value iteration and policy iteration? [Select all that apply] **1 point**

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

- [ ] Asynchronous, if it updates some states more than others.
- [ ] Synchronous, if it systematically sweeps the entire state space at each iteration.
- [ ] Asynchronous, if it does not update all states at each iteration.

6. Policy iteration and value iteration, as described in chapter four, are synchronous. **1 point**

- ◯ True
- ◯ False

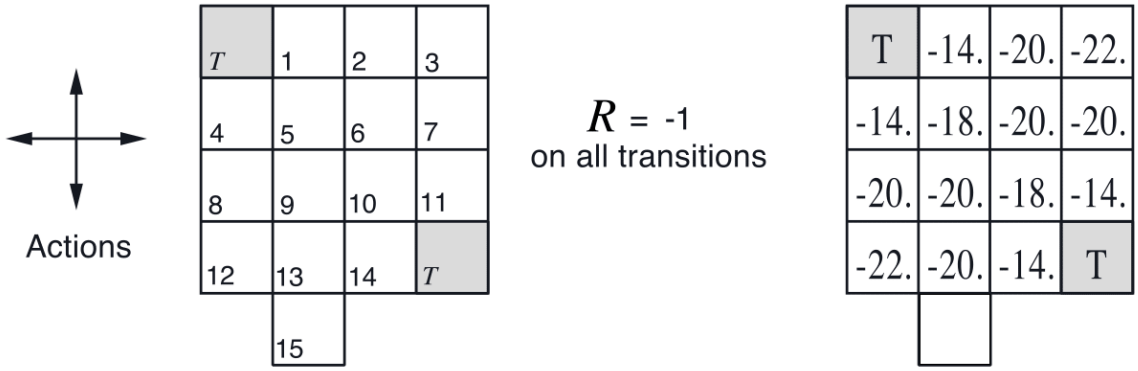
7. Which of the following is true? **1 point**

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

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

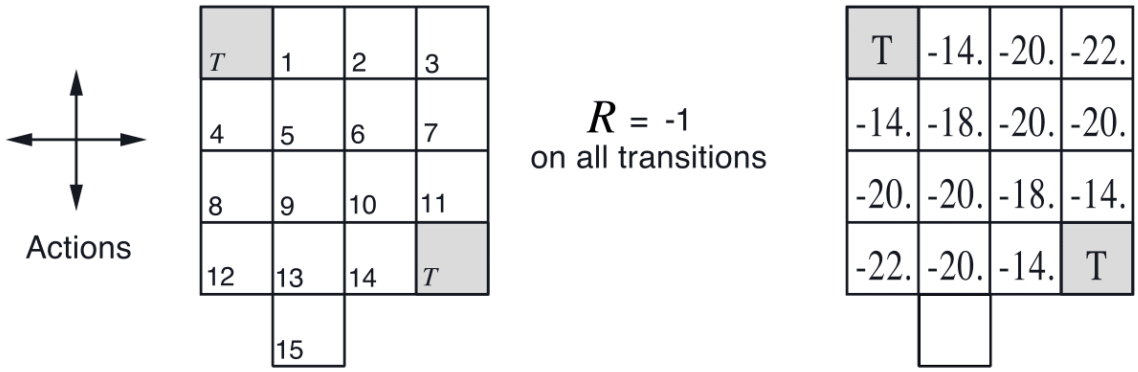
- [ ] They use a model to improve the policy.
- [ ] They compute optimal value functions.
- [ ] They learn from trial and error interaction.

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 point**



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

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 point**



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

- ○ $v(15) = -25$
- ○ $v(15) = -22$