1.

Question 1

A function which maps ___ to ___ is a value function. [Select all that apply]

State-action pairs to expected returns.

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

States to expected returns.

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

Values to states.

Values to actions.

1 / 1 point

2.

Question 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,

πleft

π

left

pi, start subscript, start text, l, e, f, t, end text, end subscript

and

πright

π

right

pi, start subscript, start text, r, i, g, h, t, end text, end subscript

. Indicate the optimal policies if

$\gamma=0$

$\gamma=0$

gamma, equals, 0

? If

$\gamma=0.9$

$\gamma=0.9$

gamma, equals, 0, point, 9

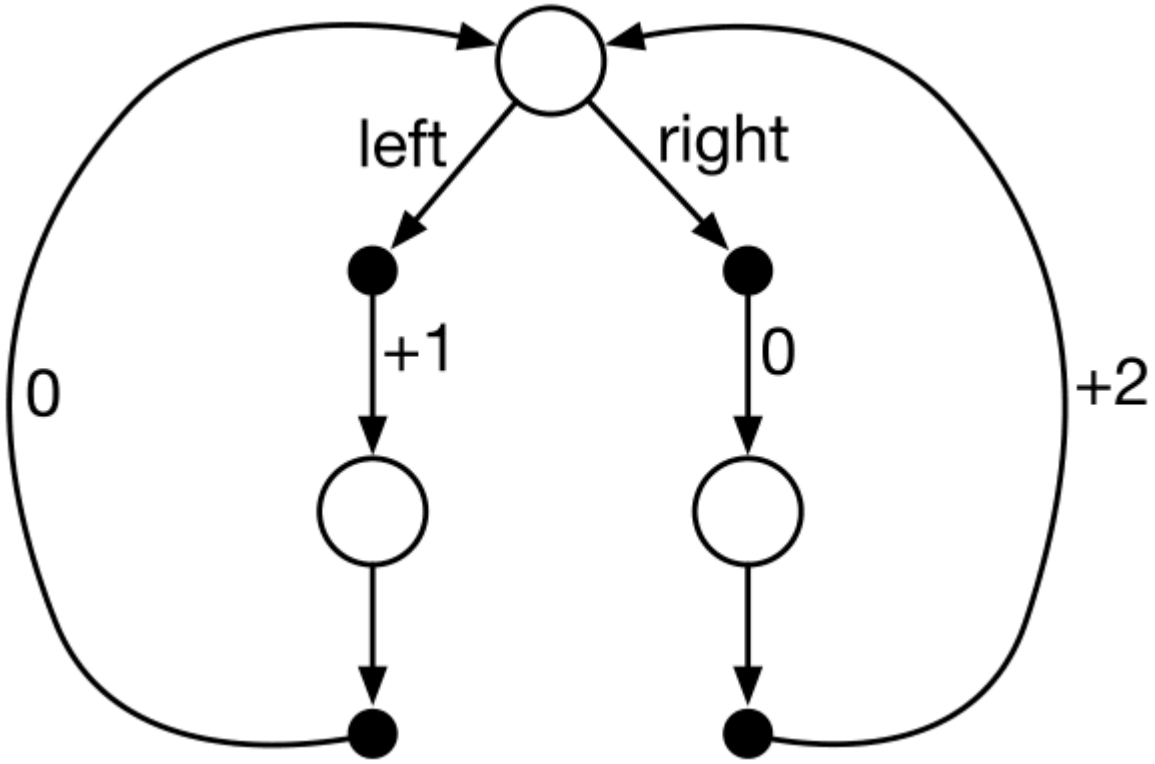
? If

$\gamma=0.5$

$\gamma=0.5$

gamma, equals, 0, point, 5

? [Select all that apply]



For

$\gamma=0.9,\pi\text{right}$

$\gamma=0.9,\pi$

right

gamma, equals, 0, point, 9, comma, pi, start subscript, start text, r, i, g, h, t, end text, end subscript

> 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.

For

$\gamma=0,\pi\text{left}$

$\gamma=0,\pi$

left

gamma, equals, 0, comma, pi, start subscript, start text, l, e, f, t, end text, end subscript

For

$\gamma=0, \pi_{\text{right}}$

$\gamma=0, \pi$

right

gamma, equals, 0, comma, pi, start subscript, start text, r, i, g, h, t, end text, end subscript

Incorrect. Take another look at the lesson: Optimal Policies.

For

$\gamma=0.9, \pi_{\text{left}}$

$\gamma=0.9,\pi$

left

gamma, equals, 0, point, 9, comma, pi, start subscript, start text, l, e, f, t, end text, end subscript

For

$\gamma=0.5,\pi$right

$\gamma=0.5,\pi$

right

gamma, equals, 0, point, 5, comma, pi, start subscript, start text, r, i, g, h, t, end text, end subscript

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

$\gamma=0.5,\pi\text{left}$

$\gamma=0.5,\pi$

left

gamma, equals, 0, point, 5, comma, pi, start subscript, start text, l, e, f, t, end text, end subscript

1 point

3.

Question 3

Every finite Markov decision process has __. [Select all that apply]

A deterministic optimal policy

Correct! Let’s say there is a policy

$\pi 1$

π

1

pi, start subscript, 1, end subscript

which does well in some states, while policy

$\pi 2$

π

2

pi, start subscript, 2, end subscript

does well in others. We could combine these policies into a third policy

$\pi 3$

π

3

pi, start subscript, 3, end subscript

, which always chooses actions according to whichever of policy

$\pi 1$

π

1

pi, start subscript, 1, end subscript

and

$\pi 2$

π

2

pi, start subscript, 2, end subscript

has the highest value in the current state.

$\pi 3$

π

3

pi, start subscript, 3, end subscript

will necessarily have a value greater than or equal to both

$\pi 1$

π

1

pi, start subscript, 1, end subscript

and

$\pi 2$

π

2

pi, start subscript, 2, end subscript

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.

A unique optimal value function

Correct! The Bellman optimality equation is actually a system of equations, one for each state, so if there are N states, then there are N equations in N unknowns. If the dynamics of the environment are known, then in principle one can solve this system of equations for the optimal value function using any one of a variety of methods for solving systems of nonlinear equations. All optimal policies share the same optimal state-value function.

A stochastic optimal policy

A unique optimal policy

1 / 1 point

4.

Question 4

The ___ of the reward for each state-action pair, the dynamics function

p

p

p

, and the policy

π

π

pi

is _____ to characterize the value function

$v\pi$

v

π

v, start subscript, pi, end subscript

. (Remember that the value of a policy

π

π

pi

at state

s

s

s

is

$$v\pi(s)=\sum a\pi(a|s)\sum s',rp(s',r|s,a)[r+\gamma v\pi(s')]$$

v

π

$$(s)=\sum$$

a

$\pi(a|s)\sum$

s

$'$

$,r$

$p(s$

$'$

$,r|s,a)[r+\gamma v$

π

$(s$

,

)]

v, start subscript, pi, end subscript, left parenthesis, s, right parenthesis, equals, sum, start subscript, a, end subscript, pi, left parenthesis, a, vertical bar, s, right parenthesis, sum, start subscript, s, prime, comma, r, end subscript, p, left parenthesis, s, prime, comma, r, vertical bar, s, comma, a, right parenthesis, open bracket, r, plus, gamma, v, start subscript, pi, end subscript, left parenthesis, s, prime, right parenthesis, close bracket

.)

Distribution; necessary

Incorrect. Take another look at the lesson: Optimal Value Functions & Bellman Optimality Equation.

Mean; sufficient

1 point

5.

Question 5

The Bellman equation for a given a policy

π

π

pi

: [Select all that apply]

Expresses state values

$v(s)$

$v(s)$

v, left parenthesis, s, right parenthesis

in terms of state values of successor states.

Correct!

Expresses the improved policy in terms of the existing policy.

Holds only when the policy is greedy with respect to the value function.

1 / 1 point

6.

Question 6

An optimal policy:

Is unique in every Markov decision process.

Is unique in every finite Markov decision process.

Is not guaranteed to be unique, even in finite Markov decision processes.

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.

1 / 1 point

7.

Question 7

The Bellman optimality equation for

$v*$

v

$*$

v, start subscript, \ast, end subscript

: [Select all that apply]

Holds for

v_π

v

π

v, start subscript, pi, end subscript

, the value function of an arbitrary policy

π

π

pi

.

Expresses the improved policy in terms of the existing policy.

Expresses state values

v*(s)

v

*

(s)

v, start subscript, \ast, end subscript, left parenthesis, s, right parenthesis

in terms of state values of successor states.

Correct!

Holds for the optimal state value function.

Correct!

Holds when the policy is greedy with respect to the value function.

1 / 1 point

8.

Question 8

Give an equation for

v_π

v_π

v, start subscript, pi, end subscript

in terms of

q_π

q_π

q, start subscript, pi, end subscript

and

π

π

pi

.

$v\pi(s)=\max_a \pi(a|s) q\pi(s,a)$

v

π

$(s)=\max$

a

$\pi(a|s)q$

π

(s,a)

v, start subscript, pi, end subscript, left parenthesis, s, right parenthesis, equals, \max, start subscript, a, end subscript, pi, left parenthesis, a, vertical bar, s, right parenthesis, q, start subscript, pi, end subscript, left parenthesis, s, comma, a, right parenthesis

$v\pi(s)=\sum a\gamma\pi(a|s)q\pi(s,a)$

v

π

$(s)=\sum$

a

$\gamma\pi(a|s)q$

π

(s,a)

v, start subscript, pi, end subscript, left parenthesis, s, right parenthesis, equals, sum, start subscript, a, end subscript, gamma, pi, left parenthesis, a, vertical bar, s, right parenthesis, q, start subscript, pi, end subscript, left parenthesis, s, comma, a, right parenthesis

$v\pi(s)=\sum a\pi(a|s)q\pi(s,a)$

v

π

$(s)=\sum$

a

$\pi(a \mid s)q$

π

(s,a)

v, start subscript, pi, end subscript, left parenthesis, s, right parenthesis, equals, sum, start subscript, a, end subscript, pi, left parenthesis, a, vertical bar, s, right

parenthesis, q, start subscript, pi, end subscript, left parenthesis, s, comma, a, right parenthesis

Correct!

vπ(s)=maxaγπ(a|s)qπ(s,a)

v

π

$(s)=\max$

a

$\gamma\pi(a\,|\,s)q$

π

(s,a)

v, start subscript, pi, end subscript, left parenthesis, s, right parenthesis, equals, \max, start subscript, a, end subscript, gamma, pi, left parenthesis, a, vertical bar, s, right parenthesis, q, start subscript, pi, end subscript, left parenthesis, s, comma, a, right parenthesis

1 / 1 point

9.

Question 9

Give an equation for

qπ

q

π

q, start subscript, pi, end subscript

in terms of

$v\pi$

v

π

v, start subscript, pi, end subscript

and the four-argument

p

p

p

.

$q\pi(s,a)=\sum s'\sum rp(s',r|s,a)\gamma[r+v\pi(s')]$

q

π

$(s,a)=\sum$

s'

$\sum$

r

$p(s',r|s,a)\gamma[r+v$

π

$(s')]$

q, start subscript, pi, end subscript, left parenthesis, s, comma, a, right parenthesis, equals, sum, start subscript, s, ', end subscript, sum, start subscript, r, end subscript, p, left parenthesis, s, ', comma, r, vertical bar, s, comma, a, right parenthesis, gamma, open bracket, r, plus, v, start subscript, pi, end subscript, left parenthesis, s, ', right parenthesis, close bracket

$q\pi(s,a)=\max s',rp(s',r|s,a)[r+\gamma v\pi(s')]$

q

π

(s,a)=max

s',r

$p(s',r|s,a)[r+\gamma v$

π

$(s')]$

q, start subscript, pi, end subscript, left parenthesis, s, comma, a, right parenthesis, equals, \max, start subscript, s, ', comma, r, end subscript, p, left parenthesis, s, ', comma, r, vertical bar, s, comma, a, right parenthesis, open bracket, r, plus, gamma, v, start subscript, pi, end subscript, left parenthesis, s, ', right parenthesis, close bracket

$$q\pi(s,a)=\sum s'\sum r p(s',r|s,a)[r+v\pi(s')]$$

q

π

$(s,a)=\sum$

s'

$\sum$

r

$p(s',r|s,a)[r+v$

π

$(s')]$

q, start subscript, pi, end subscript, left parenthesis, s, comma, a, right parenthesis, equals, sum, start subscript, s, ', end subscript, sum, start subscript, r, end subscript, p, left parenthesis, s, ', comma, r, vertical bar, s, comma, a, right parenthesis, open bracket, r, plus, v, start subscript, pi, end subscript, left parenthesis, s, ', right parenthesis, close bracket

$q\pi(s,a)=\max s',rp(s',r|s,a)\gamma[r+v\pi(s')]$

q

π

(s,a)=max

s',r

$p(s',r|s,a)\gamma[r+v$

π

$(s')]$

q, start subscript, pi, end subscript, left parenthesis, s, comma, a, right parenthesis, equals, \max, start subscript, s, ', comma, r, end subscript, p, left parenthesis, s, ', comma, r, vertical bar, s, comma, a, right parenthesis, gamma, open bracket, r, plus, v, start subscript, pi, end subscript, left parenthesis, s, ', right parenthesis, close bracket

qπ(s,a)=maxs',rp(s',r|s,a)[r+vπ(s')]

q

π

(s,a)=max

s',r

$p(s',r|s,a)[r+v$

π

$(s')]$

q, start subscript, pi, end subscript, left parenthesis, s, comma, a, right parenthesis, equals, \max, start subscript, s, ', comma, r, end subscript, p, left parenthesis, s, ', comma, r, vertical bar, s, comma, a, right parenthesis, open bracket, r, plus, v, start subscript, pi, end subscript, left parenthesis, s, ', right parenthesis, close bracket

$$q\pi(s,a)=\sum s'\sum rp(s',r|s,a)[r+\gamma v\pi(s')]$$

q

π

$(s,a)=\sum$

s'

$\sum$

r

$p(s',r|s,a)[r+\gamma v$

π

$(s')]$

q, start subscript, pi, end subscript, left parenthesis, s, comma, a, right parenthesis, equals, sum, start subscript, s, ', end subscript, sum, start subscript, r, end subscript, p, left parenthesis, s, ', comma, r, vertical bar, s, comma, a, right parenthesis, open bracket, r, plus, gamma, v, start subscript, pi, end subscript, left parenthesis, s, ', right parenthesis, close bracket

Correct!

1 / 1 point

10.

Question 10

Let

$r(s,a)$

$r(s,a)$

r, left parenthesis, s, comma, a, right parenthesis

be the expected reward for taking action

a

a

a

in state

s

s

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**]

$v\pi(s)=\sum a\pi(a|s)[r(s,a)+\gamma\sum s'p(s'|s,a)v\pi(s')]$

v

π

$(s)=\sum$

a

$\pi(a|s)[r(s,a)+\gamma\sum$

s'

$p(s'|s,a)v$

π

$(s')]$

v, start subscript, pi, end subscript, left parenthesis, s, right parenthesis, equals, sum, start subscript, a, end subscript, pi, left parenthesis, a, vertical bar, s, right parenthesis, open bracket, r, left parenthesis, s, comma, a, right parenthesis, plus, gamma, sum, start subscript, s, ', end subscript, p, left parenthesis, s, ', vertical bar, s, comma, a, right parenthesis, v, start subscript, pi, end subscript, left parenthesis, s, ', right parenthesis, close bracket

Correct!

$$q\pi(s,a)=r(s,a)+\gamma\sum s'\sum a' p(s'|s,a)\pi(a'|s')q\pi(s',a')$$

q

π

$(s,a)=r(s,a)+\gamma\sum$

s'

$\sum$

a'

$p(s'|s,a)\pi(a'|s')q$

π

(s',a')

q, start subscript, pi, end subscript, left parenthesis, s, comma, a, right parenthesis, equals, r, left parenthesis, s, comma, a, right parenthesis, plus, gamma, sum, start subscript, s, ', end subscript, sum, start subscript, a, ', end subscript, p, left parenthesis, s, ', vertical bar, s, comma, a, right parenthesis, pi, left parenthesis, a, ', vertical bar, s, ', right parenthesis, q, start subscript, pi, end subscript, left parenthesis, s, ', comma, a, ', right parenthesis

Correct!

v∗(s)=maxa[r(s,a)+γ∑s'p(s'|s,a)v∗(s')]

v

*

$(s)=\max$

a

$[r(s,a)+\gamma\sum$

s'

$p(s'|s,a)v$

*

$(s')]$

v, start subscript, \ast, end subscript, left parenthesis, s, right parenthesis, equals, \max, start subscript, a, end subscript, open bracket, r, left parenthesis, s, comma, a, right parenthesis, plus, gamma, sum, start subscript, s, ', end subscript, p, left parenthesis, s, ', vertical bar, s, comma, a, right parenthesis, v, start subscript, \ast, end subscript, left parenthesis, s, ', right parenthesis, close bracket

Correct!

$q*(s,a)=r(s,a)+\gamma\sum s' p(s' \mid s,a)\max a' q*(s',a')$

q

*

$(s,a)=r(s,a)+\gamma\sum$

s'

$p(s'|s,a)\max$

a'

q

$*$

(s',a')

q, start subscript, \ast, end subscript, left parenthesis, s, comma, a, right parenthesis, equals, r, left parenthesis, s, comma, a, right parenthesis, plus, gamma, sum, start subscript, s, ', end subscript, p, left parenthesis, s, ', vertical bar, s, comma, a, right

parenthesis, \max, start subscript, a, ’, end subscript, q, start subscript, \ast, end subscript, left parenthesis, s, ’, comma, a, ’, right parenthesis

Correct!

1 / 1 point

11.

Question 11

Consider an episodic MDP with one state and two actions (left and right). The left action has stochastic reward

1

1

1

with probability

p

p

p

and

3

3

3

with probability

$1-\mathrm{p}$

$1-p$

1, minus, p

. The right action has stochastic reward

0

0

0

with probability

q

q

q

and

10

10

10

with probability

$1-q$

$1-q$

1, minus, q

. What relationship between

p

p

p

and

q

q

q

makes the actions equally optimal?

$7+2p=10q$

$7+2p=10q$

7, plus, 2, p, equals, 10, q

Correct!

$7+2p=-10q$

$7+2p=-10q$

7, plus, 2, p, equals, minus, 10, q

13+3p=10q

$13+3p=10q$

13, plus, 3, p, equals, 10, q

7+3p=−10q

$7+3p=-10q$

7, plus, 3, p, equals, minus, 10, q

13+2p=−10q

$13+2p=-10q$

13, plus, 2, p, equals, minus, 10, q

7+3p=10q

$7+3p=10q$

7, plus, 3, p, equals, 10, q

13+2p=10q

$13+2p=10q$

13, plus, 2, p, equals, 10, q

13+3p=−10q

$13+3p=-10q$

13, plus, 3, p, equals, minus, 10, q

1 / 1 point

Like

Dislike

Report an issue