1.

Question 1

A policy is a function which maps ____ to ____.

Actions to probabilities.

States to values.

States to actions.

States to probability distributions over actions.

Correct!

Actions to probability distributions over values.

1 / 1 point

2.

Question 2

The term “backup” most closely resembles the term ___ in meaning.

Value

Update

Correct!

Diagram

1 / 1 point

3.

Question 3

At least one deterministic optimal policy exists in every Markov decision process.

True

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.

False

1 / 1 point

4.

Question 4

The optimal state-value function:

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

Is unique in every finite Markov decision process.

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.

1 / 1 point

5.

Question 5

Does adding a constant to all rewards change the set of optimal policies in episodic tasks?

Yes, adding a constant to all rewards changes the set of optimal policies.

> Correct! Adding a constant to the reward signal can make longer episodes more or less advantageous (depending on whether the constant is positive or negative).

No, as long as the relative differences between rewards remain the same, the set of optimal policies is the same.

1 / 1 point

6.

Question 6

Does adding a constant to all rewards change the set of optimal policies in continuing tasks?

Yes, adding a constant to all rewards changes the set of optimal policies.

No, as long as the relative differences between rewards remain the same, the set of optimal policies is the same.

Correct! Since the task is continuing, the agent will accumulate the same amount of extra reward independent of its behavior.

1 / 1 point

7.

Question 7

Select the equation that correctly relates

$v*$

v

$*$

v, start subscript, \ast, end subscript

to

q*

q

*

q, start subscript, \ast, end subscript

. Assume

π

π

pi

is the uniform random policy.

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

v

*

$(s)=\sum$

a,r,s'

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

*

(s')

v, start subscript, \ast, end subscript, left parenthesis, s, right parenthesis, equals, sum, start subscript, a, comma, r, comma, s, ’, end subscript, pi, left parenthesis, a, vertical bar, s, right parenthesis, p, left parenthesis, s, ’, comma, r, vertical bar, s, comma, a, right parenthesis, q, start subscript, \ast, end subscript, left parenthesis, s, ’, right parenthesis

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

v

*

$(s)=\sum$

$a,r,s’$

$\pi(a|s)p(s’,r|s,a)[r+q$

*

$(s')]$

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

v*(s)=maxaq*(s,a)

v

*

$(s)=max$

a

q

$*$

(s,a)

v, start subscript, \ast, end subscript, left parenthesis, s, right parenthesis, equals, m, a, x, start subscript, a, end subscript, q, start subscript, \ast, end subscript, left parenthesis, s, comma, a, right parenthesis

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

v

*

$(s)=\sum$

a,r,s'

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

*

$(s')]$

v, start subscript, \ast, end subscript, left parenthesis, s, right parenthesis, equals, sum, start subscript, a, comma, r, comma, s, ’, end subscript, pi, left parenthesis, a, vertical bar, s, right parenthesis, p, left parenthesis, s, ’, comma, r, vertical bar, s,

comma, a, right parenthesis, open bracket, r, plus, gamma, q, start subscript, \ast, end subscript, left parenthesis, s, ’, right parenthesis, close bracket

Incorrect. We don’t need

π

π

pi

or

p

p

p

to express

v*

v

*

v, start subscript, \ast, end subscript

in terms of

q*

q

*

q, start subscript, \ast, end subscript

. Consider that under an optimal policy, the value of a state is the same as the values of the actions supported by the policy at that state.

1 point

8.

Question 8

Select the equation that correctly relates

q*

q

*

q, start subscript, \ast, end subscript

to

v*

v

*

v, start subscript, \ast, end subscript

using four-argument function

p

p

p

.

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

q

*

$(s,a)=\sum$

s',r

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

*

$(s$

,

$)]$

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

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

q

*

$(s,a)=\sum$

s',r

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

*

$(s')]$

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

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

q

*

$(s,a)=\sum$

s',r

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

*

$(s')]$

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

Correct!

1 / 1 point

9.

Question 9

Write a policy

$\pi*$

π

$*$

pi, start subscript, \ast, end subscript

in terms of

$q*$

q

$*$

q, start subscript, \ast, end subscript

.

π*(a|s)=q*(s,a)

π

*

$(a|s)=q$

*

(s,a)

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

π∗(a|s)=maxa’q∗(s,a’)

π

∗

$(a|s)$=max

a'

q

∗

(s,a')

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

π

$*$

$(a|s)=1$ if $a=$

argmax

a

$'$

q

*

(*s*,

a

,

), else 0

Correct!

1 / 1 point

10.

Question 10

Give an equation for some

$\pi*$

π

$*$

pi, start subscript, \ast, end subscript

in terms of

$v*$

v

$*$

v, start subscript, \ast, end subscript

and the four-argument

p

p

p

.

$\pi*(a|s)=1$ if $v*(s)=\max a' \sum s', rp(s', r|s, a')[r+\gamma v*(s')]$, else 0

π

*

$(a|s)=1$ if v

*

$(s)=\max$

a'

$\sum$

s',r

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

$*$

$(s')]$, else 0

pi, start subscript, \ast, end subscript, left parenthesis, a, vertical bar, s, right parenthesis, equals, 1, start text, space, i, f, space, end text, v, start subscript, \ast, end subscript, left parenthesis, s, right parenthesis, equals, \max, start subscript, a, ’, end subscript, sum, 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, \ast, end subscript, left parenthesis, s, ’, right parenthesis, close bracket, comma, start text, space, e, l, s, e, space, end text, 0

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

π

*

$$(a|s)=\max$$

$a’$

$\sum$

s',r

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

$*$

$(s')]$

pi, start subscript, \ast, end subscript, left parenthesis, a, vertical bar, s, right parenthesis, equals, \max, start subscript, a, ', end subscript, sum, 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, \ast, end subscript, left parenthesis, s, ', right parenthesis, close bracket

$\pi*(a|s)=1$ if $v*(s)=\sum s',rp(s',r|s,a)[r+\gamma v*(s')]$, else 0

π

*

$(a|s)=1$ if v

*

$(s)=\sum$

s',r

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

*

$(s')]$, else 0

pi, start subscript, \ast, end subscript, left parenthesis, a, vertical bar, s, right parenthesis, equals, 1, start text, space, i, f, space, end text, v, start subscript, \ast, end subscript, left parenthesis, s, right parenthesis, equals, sum, 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, \ast, end subscript, left parenthesis, s, ', right parenthesis, close bracket, comma, start text, space, e, l, s, e, space, end text, 0

Correct!

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

π

*

$(a|s)=\sum$

s',r

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

*

$(s')]$

pi, start subscript, \ast, end subscript, left parenthesis, a, vertical bar, s, right parenthesis, equals, sum, 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, \ast, end subscript, left parenthesis, s, ’, right parenthesis, close bracket

1 / 1 point

Like

Dislike

Report an issue