Practice Quiz: MDPs and value iteration (not assessed)

Due No due date **Points** 8 **Questions** 6 **Time Limit** None
Allowed Attempts Unlimited

Instructions

This is short practice quiz for you to test out your understanding of the MDPs and value/policy iteration. It is advised that you read the notes and watch the videos for this unit before taking the quiz. You can take it multiple times.

Take the Quiz Again

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	74 minutes	0 out of 8

Submitted Apr 22 at 15:13

Unanswered

Question 1 **0 / 1 pts**

You want to buy a new guitar. There are three options: Maton, Fender, and Martin; but you are worried about the dreaded 'buyers remorse'.

If you buy a Maton (your dream acoustic guitar!), you think there is an 80% chance that you will feel +100 better (your reward/return); but because it is so expensive, there is a 20% chance of buyer's remorse, which will make you feel -100 (that's a *negative* reward)

If you buy a Fender, you think there is an 70% chance that you will feel +70 better; and a 30% you feel -100.

If you buy a Martin, you think there is an 60% chance that you will feel +100 better; a 20% you feel -40; and a 20% that you can sell it to your

idiot brother whose name is Martin and buys anything that bears his name, which makes you slightly happy (feel +10)

What is the expected return of the Maton?

ou Answered

orrect Answers

60 (with margin: 0)

The expected return is calculated as:

0.8 x 100 + 0.2 x -100
= 80 - 20
= 60

Unanswered

Question 2

0 / 1 pts

What is the expected return of the Fender?

ou Answered

orrect Answers

19 (with margin: 0)

The expected return is calculated as:

0.7 x 70 + 0.3 * -100
= 49 - 30
= 19

Unanswered

Question 3

0 / 1 pts

What is the expected return of the Martin?

You Answered

Correct Answers

54 (with margin: 0)

The expected return is calculated as:

0.6 x 100 + 0.2 x -40 + 0.2 * 10
= 60 - 8 + 2
= 54

Unanswered

Question 4

0 / 1 pts

Which guitar should you buy?

Correct Answer

- Maton
- Fender
- Martin

Unanswered

Question 5

0 / 3 pts

Consider the following abstract MDP with three states, s, t, and u and two actions a and b.

The transition probabilities are as follows:

P_a (t | s) = 0.6
P_a (s | s) = 0.4
P_b (u | s) = 1.0
P_b (u | t) = 1.0

Any probabilities not listed above have probability of 0.

The reward function has the following:

r(s, a, t) = 2
r(s, b, u) = 5
r(t, b, u) = 5

Assuming V(s) = V(t) = V(u) = 0, and a discount factor of 0.9, calculate the V for the first iteration to one decimal place.

V(s) =

V(t) =

V(u) =

Answer 1:

ou Answered (You left this blank)

orrect Answer 5

orrect Answer 5.0

Answer 2:

ou Answered (You left this blank)

orrect Answer 5

orrect Answer 5.0

Answer 3:

ou Answered (You left this blank)

orrect Answer 0

orrect Answer 0.0

For V(s):

Q(s, a) = P_a (t | s) * [r(s, a, t) + yV(t)] + P_a (s | s) * [r(s, a, s) + yV(s)]
= 0.6 * [2 + 0.9*0] + 0.4 * [0 + 0.9*0]
= 1.2
Q(s, b) = P_b (u | s) * [r(s, b, u) + yV(u)]
= 1.0 * [5 + 0.9*0]
= 5
max((Q(s,a), Q(s,b)) = 5
Therefore, V(s) = 5

For V(t):

Q(t, b) = P_b (u | t) * [r(t, b, u) + yV(u)]
= 1.0 * [5 + 0.9*0]
= 5
Action b is the only action, therefore V(t) = 5

For V(u):

There are no actions from u, so the value is just 0.

Unanswered

Question 6

0 / 1 pts

Take the same example from the previous question. Assume that we run value iteration until it converges and the resulting value function is:

V(s) = 12
V(t) = 10
V(u) = 0

In state s, which action should be taken: a or b?

orrect Answer

a

b

For policy extraction, we just calculate the expected reward of each action:

Q(s,a) = P_a (t | s) * [r(s, a, t) + yV(t)] + P_a (s | s) * [r(s, a, s) + yV(s)]

= 0.6 * [2 + 0.9*10] + 0.4 * [0 + 0.9*12]

= 6.6 + 4.32

= 10.92

Q(s, b) = P_b (u | s) * [r(s, b, u) + yV(u)]

= 1.0 * [5 + 0.9*0]

= 5

The argmax of these two is action a, so this is what we select.