Practice Quiz: n-step RL and MCTS (not assessed)

Due No due date **Points** 12 **Questions** 7 **Time Limit** None
Allowed Attempts Unlimited

Take the Quiz Again

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	less than 1 minute	0 out of 12

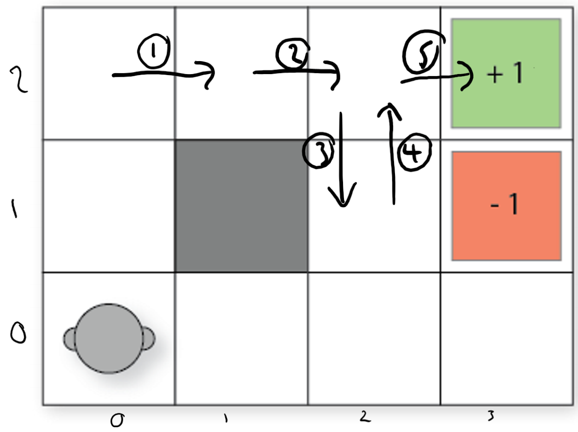
Submitted Jun 6 at 12:58

Unanswered

Question 1

0 / 3 pts

Consider the sample example from the notes:



Assuming Q(s,a)=0 for all s and a, if we (finally) traverse the episode the labelled episode, what will our Q-function look like for a 2-step update with $\alpha = 0.5$ and $\gamma = 0.9$ if we want to update the action 4?

You Answered

Correct Answers

0.405 (with margin: 0.005)

The discounted reward is $\gamma^2 = 0.9^2 = 0.81$

This update is then:

$$Q((2, 1), N) \leftarrow \$Q((1, 2), N) + \alpha[G - Q((2, 1), N)] = 0 + 0.5[0.81 - 0] = 0.405$$

Unanswered

Question 2

0 / 1 pts

Interleaved action selection (planning) and action execution is known as what?

Correct Answer

○ Online planning

○ Offline planning

○ Internet planning

○ MCTS

Unanswered

Question 3

0 / 2 pts

The four steps in each iteration of MCTS are:

1.

2.

3.

4.

Use all lower case in your answers

Answer 1:

You Answered

(You left this blank)

Correct Answer	Selection
Correct Answer	selection
Correct Answer	select
Correct Answer	selecting

Answer 2:

You Answered	(You left this blank)
Correct Answer	Expansion
Correct Answer	expansion
Correct Answer	expand
Correct Answer	expanding

Answer 3:

You Answered	(You left this blank)
Correct Answer	Simulation
Correct Answer	simulation
Correct Answer	simulating
Correct Answer	simulate

Answer 4:

You Answered	(You left this blank)
Correct Answer	Backpropagation
Correct Answer	backpropagation
Correct Answer	backpropagate
Correct Answer	backpropagating

Unanswered

Question 4

0 / 3 pts

Match the following definitions of to names of multi-armed bandit algorithms

You Answered

Exploit best action with probability 1-epsilon and random from all other actions with epsilon probability

Correct Answer epsilon-greedy

You Answered

Exploit actions proportionally based on their Q-value

Correct Answer softmax

You Answered

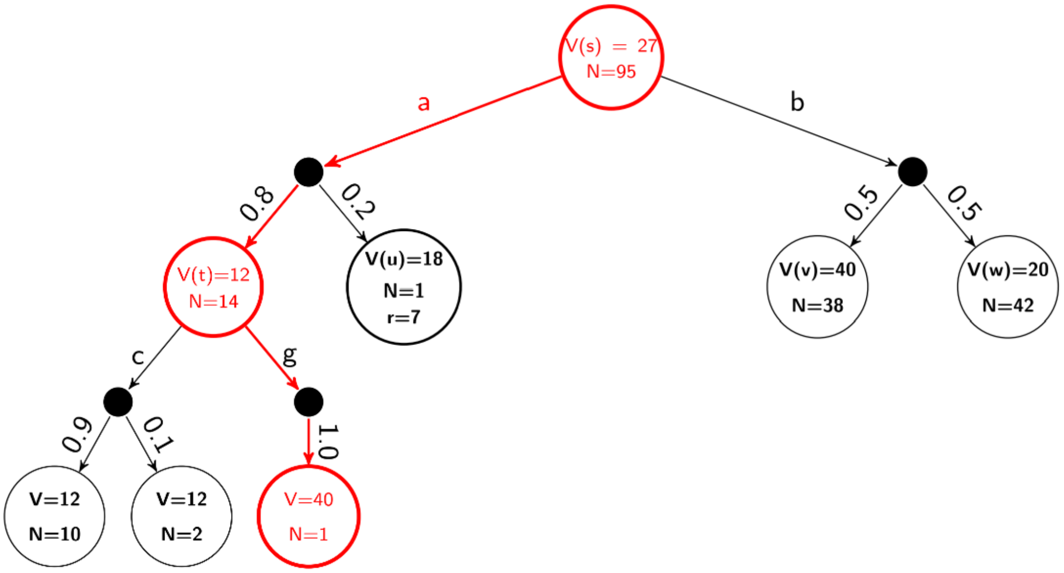
Exploit Q-value and exploit based on number of times an option has been chosen

Correct Answer UCB

Other Incorrect Match Options:

- epsilon-decreasing

The following three questions refer to the expectimax tree below:



Assume an MCTS algorithm that has just completed the steps of selection (the red path), expansion of node "t", generating with the action "g", and simulated from the new node, resulting in a value of 40 for the new node.

Perform the backpropagation step to calculate the new values for V(t) and (Vs).

Unanswered

Question 5

0 / 1 pts

Assuming $\gamma = 0.9$, what is the new value of V(t)?

ou Answered

orrect Answers

36 (with margin: 0)

$$\begin{aligned} V(t) &= \max_{a \in \{c,g\}} \textstyle\sum_{t' \in children(t)} P_a(t'|t)\ [r(t',a,t') + \gamma\ V(t')] \\ &= \max(0.9(0 + 0.9 \times 12) + 0.1(0 + 0.9 \times 12), \quad \text{(action c)} \\ &\qquad 1.0(0 + 0.9 \times 40)) \quad \text{(action g)} \\ &= \max(10.8, 36) \\ &= 36 \end{aligned}$$

Unanswered

Question 6

0 / 1 pts

Assuming $\gamma = 0.9$, what is the new value of V(s) (to one decimal place)?

You Answered

Correct Answers

30.6 (with margin: 0.1)

$$\begin{aligned} V(s) &= \max_{a\in\{a,b\}} \textstyle\sum_{s'\in children(s)} P_a(s'|s)\ [r(s,a,s') + \gamma\ V(s')] \\ &= \max(0.8(0 + 0.9 \times 36) + 0.2(7 + 0.9 \times 18), \quad \text{(action a)} \\ &\qquad\quad 0.5(0 + 0.9 \times 40) + 0.5(0 + 0.9 \times 20) \quad \text{(action b)} \\ &= \max(25.92 + 4.64,\ 18 + 9) \\ &= 30.56 \text{ rounded to } 30.6 \end{aligned}$$

Unanswered

Question 7

0 / 1 pts

Which action should you select?

Correct Answer

○ a

○ b

We know from V(s) that the maximum action is "a", so this is the one that we would select.