Q10: Potpourri

Q10: Potpourri

Part 1

0.0/2.0 points (graded)

There exists some value of $k > 0$ such that the heuristic $h(n) = k$ is admissible.

- True
- False ✔

Explanation

This heuristic is non-zero at the goal state, and so it cannot be admissible.

Submit You have used 0 of 1 attempt

Answers are displayed within the problem

Part 2

Generating Speech Output

0.0/2.0 points (graded)

A^* tree search using the heuristic $h(n) = k$ for some $k > 0$ is guaranteed to find the optimal solution.

- (•) True ✔
- ( ) False

Explanation

Each state in the fringe has priority $f(n) = g(n) + h(n) = g(n) + k$. Only the ordering of the nodes in the fringe matters, and so adding a constant c to all priorities makes no difference. So, this is equivalent to using $f(n) = g(n)$, which is UCS, which will find the optimal solution.

Submit You have used 0 of 1 attempt

Answers are displayed within the problem

Part 3

0.0/2.0 points (graded)

Consider a one-person game, where the one player's actions have non-deterministic outcomes. The player gets +1 utility for winning and -1 for losing. Mark *all* of the approaches that can be used to model and solve this game.

- [ ] Minimax with terminal values equal to +1 for wins and -1 for losses
- [x] Expectimax with terminal values equal to +1 for wins and -1 for losses ✔
- [x] Value iteration with all rewards set to 0, except wins and losses, which are set to +1 and -1 ✔
- [ ] None of the above

Generating Speech Output

Explanation
Minimax is not a good fit, because there is no minimizer - there is only a maximizer (the player) and chance nodes (the non-deterministic actions). The existence of a maximizer and chance nodes means that this is particularly suited to expectimax and value iteration.

Submit You have used 0 of 2 attempts

Answers are displayed within the problem

Pacman is offered a choice between (a) playing against 2 ghosts or (b) a lottery over playing against 0 ghosts or playing against 4 ghosts (which are equally likely). Mark the rational choice according to each utility function below; if it's a tie, mark so. Here, g is the number of ghosts Pacman has to play against.

Part 4

0.0/1.0 point (graded)
$U(g) = g$

- 2 ghosts
- lottery between 0 and 4 ghosts
- tie ✔

Explanation
For $U(g) = g$, we get $U(2) = 2$ and $U([4, 0.5; 0, 0.5]) = 0.5U(4) + 0.5U(0) = 0.5(4) + 0.5(0) = 2$. Since $2 = 2$, Pacman is indifferent.

Submit You have used 0 of 1 attempt

Generating Speech Output ...played within the problem

Part 5

0.0/1.0 point (graded)

$U(g) = -(2^g)$

- (selected) 2 ghosts ✔
- lottery between 0 and 4 ghosts
- tie

Explanation

For $U(g) = -2^g$, we get $U(2) = -2^2 = -4$ and $U([4, 0.5; 0, 0.5]) = 0.5U(4) + 0.5U(0) = 0.5(-16) + 0.5(-1) = -8.5$. Since $-4 > -8.5$, Pacman prefers the 2 ghosts.

Submit You have used 0 of 1 attempt

Answers are displayed within the problem

Part 6

0.0/1.0 point (graded)

$U(g) = 2^{(-g)} = \frac{1}{2^g}$

- 2 ghosts
- (selected) lottery between 0 and 4 ghosts ✔
- tie

Explanation

Generating Speech Output

For $U(g) = 2^{-g}$, we get $U(2) = \frac{1}{4}$ and
$U([4, 0.5; 0, 0.5]) = 0.5U(4) + 0.5U(0) = 0.5\left(\frac{1}{16}\right) + 0.5(1) = \frac{17}{32}$. Since $\frac{17}{32} > \frac{1}{4}$,
Pacman prefers the lottery.

Submit You have used 0 of 1 attempt

Answers are displayed within the problem

Part 7

0.0/1.0 point (graded)
$U(g) = 1$ if $g < 3$ else 0

- (selected) 2 ghosts ✔
- lottery between 0 and 4 ghosts
- tie

Explanation
For $U(g) = 1$ if $g < 3$ else 0, we get $U(2) = 1$ and
$U([4, 0.5; 0, 0.5]) = 0.5U(4) + 0.5U(0) = 0.5(0) + 0.5(1) = 0.5$. Since $1 > 0.5$,
Pacman prefers the 2 ghosts.

Submit You have used 0 of 1 attempt

Answers are displayed within the problem

Suppose we run value iteration in an MDP with only non-negative rewards (that is, $R(s, a, s') \geq 0$ for any (s, a, s')). Let the values on the kth iteration be $V_k(s)$ and the optimal values be $V^*(s)$. Initially, the values are 0 (that is, $V_0(s) = 0$ for any s).

Generating Speech Output

Part 8

0.0/1.0 point (graded)

Mark all of the options that are guaranteed to be true.

- [ ] For any s, a, s', $V_1(s) = R(s, a, s')$
- [ ] For any s, a, s', $V_1(s) \leq R(s, a, s')$
- [ ] For any s, a, s', $V_1(s) \geq R(s, a, s')$
- [x] None of the above are guaranteed to be true.

Explanation

$V_1(s) = \max_a \sum_{s'} T(s, a, s') R(s, a, s')$ (using the Bellman equation and setting $V_0(s') = 0$).

Now consider an MDP where the best action in state X is clockwise, which goes to state Y with a reward of 6 with probability 0.5 and goes to state Z a reward of 4 with probability 0.5. Then $V_1(X) = 0.5(6) + 0.5(4) = 5$. Notice that setting $(s, a, s') = (X, clockwise, Z)$ gives a counterexample for the second option and $(s, a, s') = (X, clockwise, Y)$ gives a counterexample for the third option.

Submit

You have used 0 of 2 attempts

Part 9

0.0/1.0 point (graded)

Mark all of the options that are guaranteed to be true.

- [ ] For any k, s, $V_k(s) = V^*(s)$
- [x] For any k, s, $V_k(s) \leq V^*(s)$ ✔
- [ ] For any k, s, $V_k(s) \geq V^*(s)$

Generating Speech Output

None of the above are guaranteed to be true.

Explanation
Intuition: Values can never decrease in an iteration. In the first iteration, since all rewards are positive, the values increase. In any other iteration, the components that contribute to $V_{k+1}(s)$ are $R(s,a,s')$ and $V(s')$. $R(s,a,s')$ is the same across all iterations, and $V(s')$ increased in the previous iteration, so we expect $V_{k+1}(s)$ to increase as well.
More formally, we can prove $V_k(s) \leq V_{k+1}(s)$ by induction.
Base Case: $V_1(s) = \max_a \sum_{s'} T(s,a,s') R(s,a,s')$.
Since $R(s,a,s') \geq 0$, $T(s,a,s') \geq 0$, we have $V_1(s) \geq 0$, and so $V_0(s) \leq V_1(s)$.
Induction: $V_{k+1}(s) = \max_a \sum_{s'} T(s,a,s') (R(s,a,s') + \gamma V_k(s'))$
$\geq \max_a \sum_{s'} T(s,a,s') (R(s,a,s') + \gamma V_{k-1}(s'))$ (using $V_k(s') \geq V_{k-1}(s')$ from the inductive hypothesis) $= V_k(s)$.
This immediately leads to $V_k(s) \leq V^*(s)$ (since we can think of $V^*(s)$ as $V_\infty(s)$).

Submit

You have used 0 of 2 attempts

Answers are displayed within the problem

Consider an arbitrary MDP where we perform Q-learning. Mark *all* of the options below in which we are guaranteed to learn the *optimal* Q-values. Assume that the learning rate α is reduced to 0 appropriately.

Part 10

0.0/2.0 points (graded)

- [ ] During learning, the agent acts according to a suboptimal policy π. The learning phase continues until convergence.
- [x] During learning, the agent chooses from the available actions at random. The learning phase continues until convergence. ✔

Generating Speech Output

- [x] During learning, in state s, the agent chooses the action a that it has chosen least often in state s, breaking ties randomly. The learning phase continues until convergence. ✔

- [ ] During learning, in state s, the agent chooses the action a that it has chosen most often in state s, breaking ties randomly. The learning phase continues until convergence.

- [ ] During learning, the agent always chooses from the available actions at random. The learning phase continues until each (s, a) pair has been seen at least 10 times.

Explanation

In order for Q-learning to converge to the *optimal* Q-values, we need every (s, a) pair to be visited infinitely often. Option 5 only does this 10 times, whereas options 1 and 4 choose only one of the many actions possible for a given state s. Only options 2 and 3 visit each (s, a) pair infinitely often.

Submit

You have used 0 of 2 attempts

Answers are displayed within the problem

Generating Speech Output