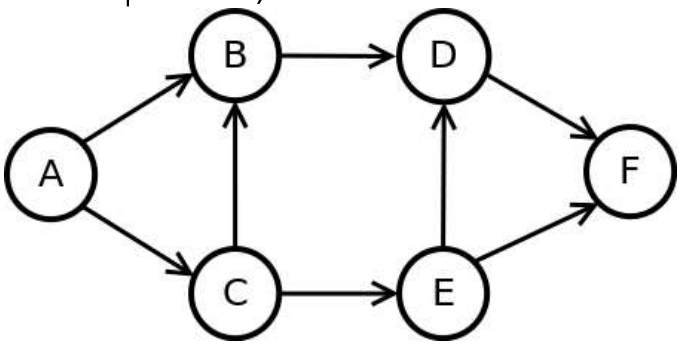
hw4_mdps_q5_value_iteration_convergence

Question 5.1: Value Iteration Convergence

0.0/4.0 points (graded)

This is a randomized question. The variables that change with each reset of the question are colored **blue**.

We will consider a simple MDP that has six states, A, B, C, D, E, and F. Each state has a single action, ***go***. An arrow from a state x to a state y indicates that it is possible to transition from state x to next state y when ***go*** is taken. If there are multiple arrows leaving a state x, transitioning to each of the next states is equally likely. The state F has no outgoing arrows: once you arrive in F, you stay in F for all future times. The reward is one for all transitions, with one exception: staying in F gets a reward of zero. Assume a discount factor = 0.5. We assume that we initialize the value of each state to 0. (Note: you should not need to explicitly run value iteration to solve this problem.)



Part 1: After how many iterations of value iteration will the value for state **F** have become exactly equal to the true optimum? (Enter inf if the values will never become equal to the true optimal but only converge to the true optimal.)

0

Answer: 0

Part 2: How many iterations of value iteration will it take for the values of all states to converge to the true optimal values? (Enter inf if the values will never become equal to the true optimal but only converge to the true optimal.)

4

Answer: 4

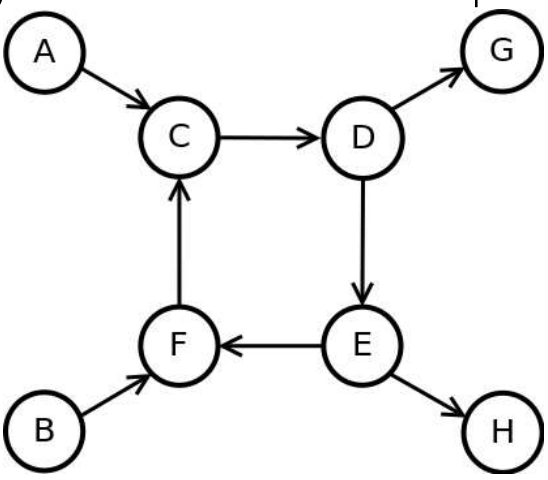
Because there are no moves from state F, we have the optimal value of F upon initializing. Since all the rewards are earned from transitions, finding the optimal value of a state amounts to finding the longest path from that state to F. For example, state D, whose longest path to F is only length 1, will find its optimal value after only one iteration.
$V^*(D) = V_1(D) = R(D, go, F) + \gamma V^*(F) = 1$
Similarly, the state A will find its optimal value after four iterations, because it will find out about its length 4 path to F after four iterations. Because A's length 4 path is the longest of the graph, it will take four iterations for all states to converge to their optimal values.

Submit

Answers are displayed within the problem

Question 5.2: Value Iteration Convergence

0.0/8.0 points (graded)
This is a randomized question. The variables that change with each reset of the question are colored **blue**.
We will consider a simple MDP that has eight states, A, B, C, D, E, F, G, and H . Each state has a single action, ***go***. An arrow from a state x to a state y indicates that it is possible to transition from state x to next state y when ***go*** is taken. If there are multiple arrows leaving a state x, transitioning to each of the next states is equally likely. The states G and H have no outgoing arrows: once you arrive in G or H, you stay in them for all future times. The reward is one for all transitions, with one exception: staying in G or H gets a reward of zero. Assume a discount factor = 0.5. We assume that we initialize the value of each state to 0. (Note: you should not need to explicitly run value iteration to solve this problem.)



Part 3: After how many iterations of value iteration will the value for state **B** have become exactly equal to the true optimum? (Enter inf if the values will never become equal to the true optimal but only converge to the true optimal.)

inf

Answer: inf

Part 4: After how many iterations of value iteration will the value for state **D** have become exactly equal to the true optimum? (Enter inf if the values will never become equal to the true optimal but only converge to the true optimal.)

inf

Answer: inf

Part 5: After how many iterations of value iteration will the value for state **G** have become exactly equal to the true optimum? (Enter inf if the values will never become equal to the true optimal but only converge to the true optimal.)

0

Answer: 0

Part 6: After how many iterations of value iteration will the value function have become exactly equal to the true optimal values? (Enter inf if the values will never become equal to the true optimal but only converge to the true optimal)

inf

Answer: inf

This MDP is very similar to the MDP in the previous question, except there is a loop. Since the rewards are only earned by transitions, the optimal policy for states A to F is to move around in the loop over and over again. Thus, those states will never become exactly equal to the optimal value. Instead, they will only converge to the optimal value. G and H will be initialized to their optimal value.

Submit

Answers are displayed within the problem