Q-Learning Homework
CS 4499/5599

Assume the deterministic 4 state world below (each cell is a state) where the immediate reward is 0 for entering all states, except the rightmost state, for which the reward is 10, and which is an absorbing state. The only actions are move right and move left (only one of which is available from the border cells). Assuming a discount factor of .8, give the final optimal Q values for each action in each state and describe an optimal policy.