Exercise 3.7

Imagine that you are designing a robot to run a maze. You decide to give it a reward of +1 for escaping from the maze and a reward of zero at all other times. The task seems to break down naturally into episodes—the successive runs through the maze—so you decide to treat it as an episodic task, where the goal is to maximize expected total reward (3.7). After running the learning agent for a while, you find that it is showing no improvement in escaping from the maze. What is going wrong? Have you effectively communicated to the agent what you want it to achieve?

Answer:

No you did not effectively tell the agent what to do because expected total reward is always 1 if you set γ to 1, and if γ is smaller than 1 it will improve extremely slowly espicially when it started far from exit. In order for it to escape from the maze faster you will need to punish it from staying in the maze and thus it will try to maximize expected total reward by find it's way out of maze as soon as possibe.

Exercise 3.8

Suppose γ = 0.5 and the following sequence of rewards is received
$R_1 = -1, R_2 = 2, R_3 = 6, R_4 = 3, R_5 = 2$, with $T = 5$. What are $G_0, G_1, \ldots, G_5$? Hint: Work backwards.

Answer

$G_5 = 0$

$G_4 = 2$

$G_3 = 4$

$G_2 = 8$

$G_1 = 6$

$G_0 = 2$

Exercise 3.9

Suppose gamma = 0.9 and the reward sequence is R1 = 2 followed by an infinite sequence of 7s. What are G1 and G0?

$G_{\ldots} = 7 + 0.9 \times 7 + 0.9^2 \times 7 +$

Saved successfully!   ✕

$= 70$

$$G_0 = 2 + 0.9 * G_1$$

$$= 65$$

Exercise 3.12

Give an equation for v_π in terms of q_π and π.

Answer:

$$V_\pi(s) = E_\pi(G_t | S_t = s) = \sum_a E_\pi(G_t | S_t = s, A_t = a) Pr(A_t = a)$$

$$= \sum_a q_\pi(s, a) \pi(a|s)$$

Exercise 3.18

$$V_\pi(s) = E_\pi(G_t | S_t = s)$$

$$V_\pi(s) = \pi(a_1|s) q_\pi(s, a_1) + \pi(a_2|s) q_\pi(s, a_2) + \pi(a_3|s) q_\pi(s, a_3)$$

$$V_\pi(s) = \sum_{i=1}^{3} \pi(a_i|s) q_\pi(s, a_i)$$

Exercise 3.19

$$q_\pi(s, a) = E[R_{t+1} + \gamma v_\pi(S_{t+1}) | S_t = s, A_t = a]$$

$$q_\pi(s, a) = \sum_{s', r} p(s', r|s, a)(r + \gamma v_\pi(s'))$$

```
###############################################################################
# Copyright (C)                                                               #
# 2016 - 2018 Shangtong Zhang(zhangshangtong.cpp@gmail.com)                   #
# 2016 Jan Hakenberg(jan.hakenberg@gmail.com)                                 #
# 2016 Tian Jun(tianjun.cpp@gmail.com)                                        #
# 2016 Kenta Shimada(hyperkentakun@gmail.com)                                 #
# Permission given to modify the code as long as you keep this                #
# declaration at the top                                                      #
###############################################################################

import numpy as np
import pickle

BOARD_ROWS = 3
BOARD_COLS = 3
BOARD_SIZE = BOARD_ROWS * BOARD_COLS
```

Saved successfully! ✕

```
        # the board is represented by an n * n array,
        # 1 represents a chessman of the player who moves first,
        # -1 represents a chessman of another player
```

```python
        # 0 represents an empty position
        self.data = np.zeros((BOARD_ROWS, BOARD_COLS))
        self.winner = None
        self.hash_val = None
        self.end = None

    # compute the hash value for one state, it's unique
    def hash(self):
        if self.hash_val is None:
            self.hash_val = 0
            for i in np.nditer(self.data):
                self.hash_val = self.hash_val * 3 + i + 1
        return self.hash_val

    # check whether a player has won the game, or it's a tie
    def is_end(self):
        if self.end is not None:
            return self.end
        results = []
        # check row
        for i in range(BOARD_ROWS):
            results.append(np.sum(self.data[i, :]))
        # check columns
        for i in range(BOARD_COLS):
            results.append(np.sum(self.data[:, i]))

        # check diagonals
        trace = 0
        reverse_trace = 0
        for i in range(BOARD_ROWS):
            trace += self.data[i, i]
            reverse_trace += self.data[i, BOARD_ROWS - 1 - i]
        results.append(trace)
        results.append(reverse_trace)

        for result in results:
            if result == 3:
                self.winner = 1
                self.end = True
                return self.end
            if result == -3:
                self.winner = -1
                self.end = True
                return self.end

        # whether it's a tie
        ____._____ = __.___(____.__(self.data))
```

Saved successfully!   ✕

```python
            self.end = True
            return self.end
```

```
            # game is still going on
            self.end = False
            return self.end


    # @symbol: 1 or -1
    # put chessman symbol in position (i, j)
    def next_state(self, i, j, symbol):
        new_state = State()
        new_state.data = np.copy(self.data)
        new_state.data[i, j] = symbol
        return new_state


    # print the board
    def print_state(self):
        for i in range(BOARD_ROWS):
            print('-------------')
            out = '| '
            for j in range(BOARD_COLS):
                if self.data[i, j] == 1:
                    token = '*'
                elif self.data[i, j] == -1:
                    token = 'x'
                else:
                    token = '0'
                out += token + ' | '
            print(out)
        print('-------------')



def get_all_states_impl(current_state, current_symbol, all_states):
    for i in range(BOARD_ROWS):
        for j in range(BOARD_COLS):
            if current_state.data[i][j] == 0:
                new_state = current_state.next_state(i, j, current_symbol)
                new_hash = new_state.hash()
                if new_hash not in all_states:
                    is_end = new_state.is_end()
                    all_states[new_hash] = (new_state, is_end)
                    if not is_end:
                        get_all_states_impl(new_state, -current_symbol, all_states)


def get_all_states():
    current_symbol = 1
    current_state = State()
    all_states = dict()
                                      = (current_state, current_state.is_end())
                                      e, current_symbol, all_states)
    return all_states
```

Saved successfully!   ✕

```python
    # all possible board configurations
    all_states = get_all_states()


class Judger:
    # @player1: the player who will move first, its chessman will be 1
    # @player2: another player with a chessman -1
    def __init__(self, player1, player2):
        self.p1 = player1
        self.p2 = player2
        self.current_player = None
        self.p1_symbol = 1
        self.p2_symbol = -1
        self.p1.set_symbol(self.p1_symbol)
        self.p2.set_symbol(self.p2_symbol)
        self.current_state = State()

    def reset(self):
        self.p1.reset()
        self.p2.reset()

    def alternate(self):
        while True:
            yield self.p1
            yield self.p2

    # @print_state: if True, print each board during the game
    def play(self, print_state=False):
        alternator = self.alternate()
        self.reset()
        current_state = State()
        self.p1.set_state(current_state)
        self.p2.set_state(current_state)
        if print_state:
            current_state.print_state()
        while True:
            player = next(alternator)
            i, j, symbol = player.act()
            next_state_hash = current_state.next_state(i, j, symbol).hash()
            current_state, is_end = all_states[next_state_hash]
            self.p1.set_state(current_state)
            self.p2.set_state(current_state)
            if print_state:
                current_state.print_state()
            if is_end:
                return current_state.winner
```

```python
class Player:
    # @step_size: the step size to update estimations
    # @epsilon: the probability to explore
```

```python
    def __init__(self, step_size=0.1, epsilon=0.1):
        # Agent's estimation of state's value
        self.estimations = dict()
        self.step_size = step_size
        # Exploit vs Explore
        self.epsilon = epsilon
        # All past states
        self.states = []
        self.greedy = []
        self.symbol = 0

    def reset(self):
        self.states = []
        self.greedy = []

    def set_state(self, state):
        self.states.append(state)
        self.greedy.append(True)

    def set_symbol(self, symbol):
        self.symbol = symbol
        for hash_val in all_states:
            state, is_end = all_states[hash_val]
            if is_end:
                if state.winner == self.symbol:
                    self.estimations[hash_val] = 1.0
                elif state.winner == 0:
                    # we need to distinguish between a tie and a lose
                    self.estimations[hash_val] = 0.5
                else:
                    self.estimations[hash_val] = 0
            else:
                self.estimations[hash_val] = 0.5

    # update value estimation
    def backup(self):
        states = [state.hash() for state in self.states]

        for i in reversed(range(len(states) - 1)):
            state = states[i]
            td_error = self.greedy[i] * (
                self.estimations[states[i + 1]] - self.estimations[state]
            )
            if td_error != 0:
              print(f"STATE[{i}]: {self.estimations[state]} => {self.estimations[state] + sel
            else:
                                         dating state value function")
                                      += self.step_size * td_error

    # choose an action based on the state
    def act(self):
        state = self.states[-1]
```

Saved successfully!　　　✕

```
        state - self.states[-1]
        next_states = []
        next_positions = []
        for i in range(BOARD_ROWS):
            for j in range(BOARD_COLS):
                if state.data[i, j] == 0:
                    next_positions.append([i, j])
                    next_states.append(state.next_state(
                        i, j, self.symbol).hash())

        # epsilon chance to explore
        if np.random.rand() < self.epsilon:
            print("Agent is taking an explore action")
            action = next_positions[np.random.randint(len(next_positions))]
            action.append(self.symbol)
            self.greedy[-1] = False
            return action
        # print("Agent is taking an exploit action")
        values = []
        for hash_val, pos in zip(next_states, next_positions):
            values.append((self.estimations[hash_val], pos))
        # to select one of the actions of equal value at random due to Python's sort is stabl
        np.random.shuffle(values)
        values.sort(key=lambda x: x[0], reverse=True)
        action = values[0][1]
        action.append(self.symbol)

        return action

    def save_policy(self):
        with open('policy_%s.bin' % ('first' if self.symbol == 1 else 'second'), 'wb') as f:
            pickle.dump(self.estimations, f)

    def load_policy(self):
        with open('policy_%s.bin' % ('first' if self.symbol == 1 else 'second'), 'rb') as f:
            self.estimations = pickle.load(f)
```

Comment

1. An agent should be able to receive information on current state, and it should have state value function and action value function, since each action will only lead to one certain state and this is a greedy algorithm that always maximize immediate returns, thus action value function will be next state's value function. we can see in act() function action is selected based on

Saved successfully! ✕

episilon value, when never a random number fall under the episilon value it will explore by randomly selecting one of the next positions. And in backup()

function we see how the state value function is updated, if it reaches a state that game ends, it sets the state's value based on game result: win = 1, tie = 0.5 and lose = 0. Then it will update all the states that leads toward the win backward, and it will only update if it's an exploit action.

```python
# human interface
# input a number to put a chessman
# | q | w | e |
# | a | s | d |
# | z | x | c |
class HumanPlayer:
    def __init__(self, **kwargs):
        self.symbol = None
        self.keys = ['q', 'w', 'e', 'a', 's', 'd', 'z', 'x', 'c']
        self.state = None

    def reset(self):
        pass

    def set_state(self, state):
        self.state = state

    def set_symbol(self, symbol):
        self.symbol = symbol

    def act(self):
        self.state.print_state()
        key = input("Input your position:")
        data = self.keys.index(key)
        i = data // BOARD_COLS
        j = data % BOARD_COLS
        return i, j, self.symbol


def train(epochs, print_every_n=500):
    player1 = Player(epsilon=0.01)
    player2 = Player(epsilon=0.01)
    judger = Judger(player1, player2)
    player1_win = 0.0
    player2_win = 0.0
    for i in range(1, epochs + 1):
        winner = judger.play(print_state=False)
        if winner == 1:
            player1_win += 1
```

```python
        print('Epoch %d, player 1 winrate: %.02f, player 2 winrate: %.02f' % (i, player1_
        player1.backup()
        player2.backup()
```

```
        player 2.backup()
        judger.reset()
    player1.save_policy()
    player2.save_policy()


def compete(turns):
    player1 = Player(epsilon=0)
    player2 = Player(epsilon=0)
    judger = Judger(player1, player2)
    player1.load_policy()
    player2.load_policy()
    player1_win = 0.0
    player2_win = 0.0
    for _ in range(turns):
        winner = judger.play()
        if winner == 1:
            player1_win += 1
        if winner == -1:
            player2_win += 1
        judger.reset()
    print('%d turns, player 1 win %.02f, player 2 win %.02f' % (turns, player1_win / turns, p


# The game is a zero sum game. If both players are playing with an optimal strategy, every ga
# So we test whether the AI can guarantee at least a tie if it goes second.
def play():
    while True:
        player1 = HumanPlayer()
        player2 = Player(epsilon=0)
        judger = Judger(player1, player2)
        player2.load_policy()
        winner = judger.play()
        if winner == player2.symbol:
            print("You lose!")
        elif winner == player1.symbol:
            print("You win!")
        else:
            print("It is a tie!")


if __name__ == '__main__':
    train(int(1e2))
    compete(int(1e1))
    play()
```

Saved successfully!   ✕

```
STATE[7]: 0.5 => 0.45
STATE[6]: 0.5 => 0.495
STATE[5]: 0.5 => 0.4995
STATE[4]: 0.5 => 0.49995
STATE[3]: 0.5 => 0.499995
STATE[2]: 0.5 => 0.4999995
STATE[1]: 0.5 => 0.49999994999999997
STATE[0]: 0.5 => 0.499999995
STATE[7]: 0.5 => 0.55
STATE[6]: 0.5 => 0.505
STATE[5]: 0.5 => 0.5005
STATE[4]: 0.5 => 0.50005
STATE[3]: 0.5 => 0.500005
STATE[2]: 0.5 => 0.5000005
STATE[1]: 0.5 => 0.50000005
STATE[0]: 0.5 => 0.500000005
STATE[4]: 0.5 => 0.55
STATE[3]: 0.5 => 0.505
STATE[2]: 0.5 => 0.5005
STATE[1]: 0.5 => 0.50005
STATE[0]: 0.499999995 => 0.5000049955
STATE[4]: 0.5 => 0.45
STATE[3]: 0.5 => 0.495
STATE[2]: 0.5 => 0.4995
STATE[1]: 0.5 => 0.49995
STATE[0]: 0.500000005 => 0.49999500449999995
Agent is taking an explore action
STATE[6]: 0.5 => 0.55
STATE[5]: 0.5 => 0.505
STATE[4]: 0.5 => 0.5005
STATE[3]: 0.5 => 0.50005
STATE[2]: 0.5 => 0.500005
STATE[1]: 0.5 => 0.5000005
Explore, not updating state value function
STATE[6]: 0.5 => 0.45
STATE[5]: 0.5 => 0.495
STATE[4]: 0.5 => 0.4995
STATE[3]: 0.5 => 0.49995
STATE[2]: 0.5 => 0.499995
STATE[1]: 0.5 => 0.4999995
STATE[0]: 0.49999500449999995 => 0.49999545404999995
STATE[6]: 0.5 => 0.55
STATE[5]: 0.5 => 0.505
STATE[4]: 0.5 => 0.5005
STATE[3]: 0.5 => 0.50005
STATE[2]: 0.5 => 0.500005
STATE[1]: 0.50005 => 0.5000455
STATE[0]: 0.5000049955 => 0.50000904595
STATE[6]: 0.5 => 0.45
STATE[5]: 0.5 => 0.495
STATE[4]: 0.5 => 0.4995
```

Saved successfully!　　　　　　　✕

```
STATE[0]: 0.49999545404999995 => 0.49999135864499994
STATE[4]: 0.5 => 0.55
STATE[3]: 0.5 => 0.505
```

```
STATE[2]: 0.5 => 0.5005
STATE[1]: 0.5000455 => 0.50009095
STATE[0]: 0.50000904595 => 0.500017236355
STATE[4]: 0.5 => 0.45
STATE[3]: 0.5 => 0.495
STATE[2]: 0.5 => 0.4995
STATE[1]: 0.4999545 => 0.49990905
STATE[0]: 0.49999135864499994 => 0.49998312778049997
STATE[5]: 0.5 => 0.45
STATE[4]: 0.5 => 0.495
STATE[3]: 0.5 => 0.4995
STATE[2]: 0.5 => 0.49995
STATE[1]: 0.50009095 => 0.500076855
STATE[0]: 0.500017236355 => 0.5000231982194999
STATE[5]: 0.5 => 0.55
STATE[4]: 0.5 => 0.505
STATE[3]: 0.5 => 0.5005
STATE[2]: 0.5 => 0.50005
STATE[1]: 0.49990905 => 0.499923145
STATE[0]: 0.49998312778049997 => 0.49997712950245
STATE[4]: 0.5 => 0.55
STATE[3]: 0.5 => 0.505
STATE[2]: 0.49995 => 0.500455
STATE[1]: 0.500076855 => 0.5001146694999999
STATE[0]: 0.5000231982194999 => 0.5000323453475499
STATE[4]: 0.5 => 0.45
STATE[3]: 0.5 => 0.495
STATE[2]: 0.50005 => 0.499545
STATE[1]: 0.499923145 => 0.4998853305
STATE[0]: 0.49997712950245 => 0.499967949602205
Agent is taking an explore action
STATE[6]: 0.5 => 0.55
STATE[5]: 0.5 => 0.505
STATE[4]: 0.5 => 0.5005
STATE[3]: 0.50005 => 0.500095
STATE[2]: 0.500005 => 0.5000140000000001
STATE[1]: 0.5001146694999999 => 0.50010460255
STATE[0]: 0.5000323453475499 => 0.500039571067795
STATE[6]: 0.5 => 0.45
STATE[5]: 0.5 => 0.495
STATE[4]: 0.5 => 0.4995
STATE[3]: 0.49995 => 0.499905
STATE[2]: 0.499995 => 0.4999860000000004
Explore, not updating state value function
STATE[0]: 0.499967949602205 => 0.49995968769198446
STATE[7]: 0.5 => 0.45
STATE[6]: 0.5 => 0.495
STATE[5]: 0.5 => 0.4995
STATE[4]: 0.5 => 0.49995
STATE[3]: 0.5 => 0.499995
STATE[2]: 0.5 => 0.4999995
STATE[1]: 0.50010460255 => 0.500094092295
                            0.5000450231905155
STATE[5]: 0.5 => 0.5005
STATE[4]: 0.5 => 0.50005
STATE[3]: 0.5 => 0.500005
```

Saved successfully! ✕

```
STATE[2]: 0.5 => 0.5000005
STATE[1]: 0.4998853305 => 0.49989684745
STATE[0]: 0.49995968769198446 => 0.499953403667786
STATE[4]: 0.5 => 0.55
STATE[3]: 0.5 => 0.505
STATE[2]: 0.4999995 => 0.50049955
STATE[1]: 0.500094092295 => 0.5001346380655001
STATE[0]: 0.5000450231905155 => 0.5000539846780139
STATE[4]: 0.5 => 0.45
STATE[3]: 0.5 => 0.495
STATE[2]: 0.5000005 => 0.49950045
STATE[1]: 0.49989684745 => 0.49985720770500003
STATE[0]: 0.499953403667786 => 0.4999437840715074
STATE[5]: 0.5 => 0.45
STATE[4]: 0.495 => 0.4905
STATE[3]: 0.5 => 0.49905
STATE[2]: 0.5 => 0.499905
STATE[1]: 0.5001346380655001 => 0.5001116742589501
STATE[0]: 0.5000539846780139 => 0.5000597536361076
STATE[5]: 0.5 => 0.55
STATE[4]: 0.505 => 0.5095000000000001
STATE[3]: 0.5 => 0.50095
STATE[2]: 0.5 => 0.500095
STATE[1]: 0.49985720770500003 => 0.4998809869345
STATE[0]: 0.4999437840715074 => 0.49993750435780665
STATE[4]: 0.5 => 0.55
STATE[3]: 0.5 => 0.505
STATE[2]: 0.499905 => 0.5004145
STATE[1]: 0.5001116742589501 => 0.5001419568330551
STATE[0]: 0.5000597536361076 => 0.5000679739558023
STATE[4]: 0.5 => 0.45
STATE[3]: 0.5 => 0.495
STATE[2]: 0.500095 => 0.49958549999999996
STATE[1]: 0.4998809869345 => 0.49985143824105
STATE[0]: 0.49993750435780665 => 0.499928897746131
STATE[4]: 0.5 => 0.55
STATE[3]: 0.5 => 0.505
STATE[2]: 0.5 => 0.5005
STATE[1]: 0.5001419568330551 => 0.5001777611497495
STATE[0]: 0.5000679739558023 => 0.500078952675197
STATE[4]: 0.5 => 0.45
STATE[3]: 0.5 => 0.495
STATE[2]: 0.5 => 0.4995
STATE[1]: 0.49985143824105 => 0.499816294416945
STATE[0]: 0.499928897746131 => 0.4999176374132124
STATE[5]: 0.5 => 0.45
STATE[4]: 0.5 => 0.495
STATE[3]: 0.5 => 0.4995
STATE[2]: 0.5 => 0.49995
STATE[1]: 0.5001777611497495 => 0.5001549850347746
STATE[0]: 0.500078952675197 => 0.5000865559111547
STATE[5]: 0.5 => 0.55
```

Saved successfully!   ✕

```
STATE[2]: 0.5 => 0.50005
STATE[1]: 0.499816294416945 => 0.49983966497525045
STATE[0]: 0.4999176374132124 => 0.4999098401694162
Agent is taking an explore action
```

```
Agent is taking an explore action
Explore, not updating state value function
Explore, not updating state value function
STATE[2]: 0.49995 => 0.499955
STATE[1]: 0.5001549850347746 => 0.5001349865312972
STATE[0]: 0.5000865559111547 => 0.5000913989731689
STATE[4]: 0.5 => 0.45
STATE[3]: 0.5 => 0.495
STATE[2]: 0.50005 => 0.499545
STATE[1]: 0.49983966497525045 => 0.4998101984777254
STATE[0]: 0.4999098401694162 => 0.4998998760002471
STATE[8]: 0.5 => 0.55
STATE[7]: 0.5 => 0.505
STATE[6]: 0.5 => 0.5005
STATE[5]: 0.5 => 0.50005
STATE[4]: 0.5 => 0.500005
STATE[3]: 0.500095 => 0.5000859999999999
STATE[2]: 0.5000140000000001 => 0.5000212
STATE[1]: 0.5001349865312972 => 0.5001236078781675
STATE[0]: 0.5000913989731689 => 0.5000946198636688
STATE[8]: 0.5 => 0.45
STATE[7]: 0.5 => 0.495
STATE[6]: 0.5 => 0.4995
STATE[5]: 0.5 => 0.49995
STATE[4]: 0.5 => 0.499995
STATE[3]: 0.499905 => 0.49991399999999997
STATE[2]: 0.49998600000000004 => 0.49997880000000006
STATE[1]: 0.4998101984777254 => 0.49982705862995286
STATE[0]: 0.4998998760002471 => 0.4998925942632177
STATE[6]: 0.5 => 0.55
STATE[5]: 0.5 => 0.505
STATE[4]: 0.5 => 0.5005
STATE[3]: 0.5000859999999999 => 0.5001273999999999
STATE[2]: 0.5000212 => 0.50003182
STATE[1]: 0.5001236078781675 => 0.5001144290903508
STATE[0]: 0.5000946198636688 => 0.500096600786337
STATE[6]: 0.5 => 0.45
STATE[5]: 0.5 => 0.495
STATE[4]: 0.5 => 0.4995
STATE[3]: 0.49991399999999997 => 0.4998726
STATE[2]: 0.49997880000000006 => 0.49996818000000004
STATE[1]: 0.49982705862995286 => 0.49984117076695755
STATE[0]: 0.4998925942632177 => 0.4998874519135917
Explore, not updating state value function
Explore, not updating state value function
Explore, not updating state value function
Explore, not updating state value function
Explore, not updating state value function
STATE[3]: 0.5001273999999999 => 0.5001146599999999
STATE[2]: 0.50003182 => 0.500040104
STATE[1]: 0.5001144290903508 => 0.5001069965813156
STATE[0]: 0.500096600786337 => 0.5000976403658348
                                ue function
Saved successfully!            ×  ue function
                                ue function
Explore, not updating state value function
Explore, not updating state value function
STATE[3]: 0.4998726 => 0.49988534
```

```
STATE[2]: 0.49996818000000004 => 0.499959896
STATE[1]: 0.49984117076695755 => 0.4998530432902618
STATE[0]: 0.4998874519135917 => 0.4998840110512587
STATE[8]: 0.5 => 0.55
STATE[7]: 0.5 => 0.505
STATE[6]: 0.5 => 0.5005
STATE[5]: 0.5 => 0.50005
STATE[4]: 0.5 => 0.500005
STATE[3]: 0.5001146599999999 => 0.500103694
STATE[2]: 0.500040104 => 0.500046463
STATE[1]: 0.5001069965813156 => 0.5001009432231841
STATE[0]: 0.5000976403658348 => 0.5000979706515698
STATE[8]: 0.5 => 0.45
STATE[7]: 0.5 => 0.495
STATE[6]: 0.5 => 0.4995
STATE[5]: 0.5 => 0.49995
STATE[4]: 0.5 => 0.499995
STATE[3]: 0.49988534 => 0.499896306
STATE[2]: 0.499959896 => 0.499953537
STATE[1]: 0.4998530432902618 => 0.49986309266123563
STATE[0]: 0.4998840110512587 => 0.4998819192122564
STATE[6]: 0.5 => 0.55
STATE[5]: 0.5 => 0.505
STATE[4]: 0.5 => 0.5005
STATE[3]: 0.500103694 => 0.5001433246
STATE[2]: 0.500046463 => 0.50005614916
STATE[1]: 0.5001009432231841 => 0.5000964638168657
STATE[0]: 0.5000979706515698 => 0.5000978199680994
STATE[6]: 0.5 => 0.45
STATE[5]: 0.5 => 0.495
STATE[4]: 0.5 => 0.4995
STATE[3]: 0.499896306 => 0.4998566754
STATE[2]: 0.499953537 => 0.49994385084
STATE[1]: 0.49986309266123563 => 0.49987116847911206
STATE[0]: 0.4998819192122564 => 0.49988084413894196
STATE[7]: 0.5 => 0.45
STATE[6]: 0.5 => 0.495
STATE[5]: 0.50005 => 0.499545
STATE[4]: 0.500005 => 0.49995900000000004
STATE[3]: 0.5001433246 => 0.50012489214
STATE[2]: 0.50005614916 => 0.5000630234579999
STATE[1]: 0.5000964638168657 => 0.5000931197809791
STATE[0]: 0.5000978199680994 => 0.5000973499493874
STATE[7]: 0.5 => 0.55
STATE[6]: 0.5 => 0.505
STATE[5]: 0.49995 => 0.500455
STATE[4]: 0.499995 => 0.5000410000000001
STATE[3]: 0.4998566754 => 0.49987510786
STATE[2]: 0.49994385084 => 0.499936976542
STATE[1]: 0.49987116847911206 => 0.49987774928540085
STATE[0]: 0.49988084413894196 => 0.49988053465358784
STATE[6]: 0.5 => 0.55
```
```
                              > 0.5004631
STATE[3]: 0.50012489214 => 0.500158712926
STATE[2]: 0.5000630234579999 => 0.5000725924048
STATE[1]: 0.5000931197809791 => 0.5000910670433613
STATE[0]: 0.5000973499493874 => 0.5000967216587847
```

Saved successfully!   ✕

```
STATE[6]: 0.5 => 0.45
STATE[5]: 0.5 => 0.495
STATE[4]: 0.500041000000001 => 0.49953690000000006
STATE[3]: 0.49987510786 => 0.499841287074
STATE[2]: 0.499936976542 => 0.4999274075952
STATE[1]: 0.49987774928540085 => 0.49988271511638077
STATE[0]: 0.49988053465358784 => 0.49988075269986715
Explore, not updating state value function
Explore, not updating state value function
Explore, not updating state value function
STATE[5]: 0.50005 => 0.500045
STATE[4]: 0.500005 => 0.500009
STATE[3]: 0.500158712926 => 0.5001437416334
STATE[2]: 0.5000725924048 => 0.5000797073276599
STATE[1]: 0.5000910670433613 => 0.5000899310717911
STATE[0]: 0.5000967216587847 => 0.5000960426000853
Explore, not updating state value function
Explore, not updating state value function
Explore, not updating state value function
STATE[5]: 0.49995 => 0.499955
STATE[4]: 0.499995 => 0.499991
STATE[3]: 0.499841287074 => 0.49985625836659997
STATE[2]: 0.4999274075952 => 0.49992029267234
STATE[1]: 0.49988271511638077 => 0.4998864728719767
STATE[0]: 0.49988075269986715 => 0.4998813247170781
Explore, not updating state value function
Explore, not updating state value function
Explore, not updating state value function
STATE[5]: 0.500045 => 0.5000405
STATE[4]: 0.500009 => 0.50001215
STATE[3]: 0.5001437416334 => 0.50013058247006
STATE[2]: 0.5000797073276599 => 0.5000847948418999
STATE[1]: 0.5000899310717911 => 0.500089417448802
STATE[0]: 0.5000960426000853 => 0.500095380084957
Explore, not updating state value function
Explore, not updating state value function
Explore, not updating state value function
STATE[5]: 0.499955 => 0.4999595
STATE[4]: 0.499991 => 0.49998785
STATE[3]: 0.49985625836659997 => 0.49986941752993996
STATE[2]: 0.49992029267234 => 0.4999152051581
STATE[1]: 0.4998864728719767 => 0.499889346100589
STATE[0]: 0.4998813247170781 => 0.49988212685542915
Explore, not updating state value function
Explore, not updating state value function
Explore, not updating state value function
STATE[5]: 0.5000405 => 0.50003645
STATE[4]: 0.50001215 => 0.50001458
STATE[3]: 0.50013058247006 => 0.5001189822230541
STATE[2]: 0.5000847948418999 => 0.5000882135800153
STATE[1]: 0.500089417448802 => 0.5000892970619233
STATE[0]: 0.500095380084957 => 0.5000947717826536
                              ue function
                              ue function
Explore, not updating state value function
STATE[5]: 0.4999595 => 0.49996355
STATE[4]: 0.49998785 => 0.49998542
STATE[3]: 0.49986941752993996 => 0.499881017776946
```

Saved successfully! ✕

STATE[3]: 0.49986941752993996 => 0.4998810177769a8
STATE[2]: 0.4999152051581 => 0.4999117864199846
STATE[1]: 0.499889346100589 => 0.49989159013252854
STATE[0]: 0.49988212685542915 => 0.49988307318313907
STATE[7]: 0.5 => 0.45
STATE[6]: 0.5 => 0.495
STATE[5]: 0.50003645 => 0.49953280499999997
STATE[4]: 0.50001458 => 0.49996640249999996
STATE[3]: 0.5001189822230541 => 0.5001037242507487
STATE[2]: 0.5000882135800153 => 0.5000897646470887
STATE[1]: 0.5000892970619233 => 0.5000893438204398
STATE[0]: 0.5000947717826536 => 0.5000942289864322
STATE[7]: 0.5 => 0.55
STATE[6]: 0.5 => 0.505
STATE[5]: 0.49996355 => 0.500467195
STATE[4]: 0.49998542 => 0.5000335975
STATE[3]: 0.499881017776946 => 0.4998962757492514
STATE[2]: 0.4999117864199846 => 0.4999102353529113
STATE[1]: 0.49989159013252854 => 0.4998934546545668
STATE[0]: 0.49988307318313907 => 0.4998841113302818
STATE[6]: 0.5 => 0.55
STATE[5]: 0.5 => 0.505
STATE[4]: 0.49996640249999996 => 0.50046976225
STATE[3]: 0.5001037242507487 => 0.5001403280506738
STATE[2]: 0.5000897646470887 => 0.5000948209874472
STATE[1]: 0.5000893438204398 => 0.5000898915371406
STATE[0]: 0.5000942289864322 => 0.500093795241503
STATE[6]: 0.5 => 0.45
STATE[5]: 0.5 => 0.495
STATE[4]: 0.5000335975 => 0.49953023775
STATE[3]: 0.4998962757492514 => 0.49985967194932623
STATE[2]: 0.4999102353529113 => 0.49990517901255277
STATE[1]: 0.4998934546545668 => 0.4998946270903654
STATE[0]: 0.4998841113302818 => 0.4998851629062902
STATE[6]: 0.5 => 0.55
STATE[5]: 0.505 => 0.5095000000000001
STATE[4]: 0.5004631 => 0.5013667900000001
STATE[3]: 0.5001403280506738 => 0.5002629742456064
STATE[2]: 0.5000948209874472 => 0.500111636313263
STATE[1]: 0.5000898915371406 => 0.5000920660147528
STATE[0]: 0.500093795241503 => 0.5000936223188279
STATE[6]: 0.5 => 0.45
STATE[5]: 0.495 => 0.4905
STATE[4]: 0.4995369000000006 => 0.49863321000000005
STATE[3]: 0.49985967194932623 => 0.4997370257543936
STATE[2]: 0.49990517901255277 => 0.49988836368673684
STATE[1]: 0.4998946270903654 => 0.49989400075000257
STATE[0]: 0.4998851629062902 => 0.49988604669066145
STATE[6]: 0.5 => 0.55
STATE[5]: 0.505 => 0.5095000000000001
STATE[4]: 0.50046976225 => 0.5013727860249999
STATE[3]: 0.5002629742456064 => 0.5003739554235458
                                  0.5001378682242913
Saved successfully!   ✕   0.5000966462357066
                                  0.5000939247105158
STATE[6]: 0.5 => 0.45
STATE[5]: 0.495 => 0.4905
STATE[4]: 0.49953023775 => 0.498627213975

```
STATE[3]: 0.4997370257543936 => 0.49962604457645426
STATE[2]: 0.49988836368673684 => 0.4998621317757086
STATE[1]: 0.49989400075000257 => 0.49989081385257317
STATE[0]: 0.49988604669066145 => 0.49988652340685263
Explore, not updating state value function
Explore, not updating state value function
Explore, not updating state value function
STATE[5]: 0.505 => 0.5045
STATE[4]: 0.5005 => 0.5008999999999999
STATE[3]: 0.5003739554235458 => 0.5004265598811912
STATE[2]: 0.5001378682242913 => 0.5001667373899813
STATE[1]: 0.5000966462357066 => 0.5001036553511341
STATE[0]: 0.5000939247105158 => 0.5000948977745776
Explore, not updating state value function
Explore, not updating state value function
Explore, not updating state value function
STATE[5]: 0.495 => 0.4955
STATE[4]: 0.4995 => 0.4991
STATE[3]: 0.49962604457645426 => 0.49957344011880883
STATE[2]: 0.4998621317757086 => 0.4998332626100186
STATE[1]: 0.49989081385257317 => 0.49988505872831773
STATE[0]: 0.49988652340685263 => 0.4998863769389991
STATE[6]: 0.5 => 0.55
STATE[5]: 0.505 => 0.5095000000000001
STATE[4]: 0.5005 => 0.5014
STATE[3]: 0.5004265598811912 => 0.5005239038930721
STATE[2]: 0.5001667373899813 => 0.5002024540402904
STATE[1]: 0.5001036553511341 => 0.5001135352200498
STATE[0]: 0.5000948977745776 => 0.5000967615191249
STATE[6]: 0.5 => 0.45
STATE[5]: 0.495 => 0.4905
STATE[4]: 0.4995 => 0.4986
STATE[3]: 0.49957344011880883 => 0.49947609610692795
STATE[2]: 0.4998332626100186 => 0.49979754595970954
STATE[1]: 0.49988505872831773 => 0.4998763074514569
STATE[0]: 0.4998863769389991 => 0.4998853699902449
STATE[6]: 0.5 => 0.55
STATE[5]: 0.505 => 0.5095000000000001
STATE[4]: 0.5005 => 0.5014
STATE[3]: 0.5005239038930721 => 0.5006115135037649
STATE[2]: 0.5002024540402904 => 0.5002433599866378
STATE[1]: 0.5001135352200498 => 0.5001265176967086
STATE[0]: 0.5000967615191249 => 0.5000997371368833
STATE[6]: 0.5 => 0.45
STATE[5]: 0.495 => 0.4905
STATE[4]: 0.4995 => 0.4986
STATE[3]: 0.49947609610692795 => 0.49938848649623513
STATE[2]: 0.49979754595970954 => 0.4997566400133621
STATE[1]: 0.4998763074514569 => 0.4998643407076474
STATE[0]: 0.4998853699902449 => 0.49988326706198516
STATE[6]: 0.5 => 0.55
STATE[5]: 0.505 => 0.5095000000000001
```

Saved successfully!   ×
```
                              0.5006903621533884
STATE[2]: 0.5002433599866378 => 0.50028806602033128
STATE[1]: 0.5001265176967086 => 0.5001426719473689
STATE[0]: 0.5000997371368833 => 0.5001040306179318
STATE[6]: 0.5 => 0.45
```

```
STATE[5]: 0.495 => 0.4905
STATE[4]: 0.4995 => 0.4986
STATE[3]: 0.49938848649623513 => 0.49930963784661164
STATE[2]: 0.4997566400133621 => 0.49971193979668704
STATE[1]: 0.4998643407076474 => 0.49984910061655136
STATE[0]: 0.49988326706198516 => 0.49987985041744176
Explore, not updating state value function
Explore, not updating state value function
Explore, not updating state value function
STATE[5]: 0.5045 => 0.50405
STATE[4]: 0.5008999999999999 => 0.501215
STATE[3]: 0.5006903621533884 => 0.5007428259380495
STATE[2]: 0.5002880602033128 => 0.5003335367767865
STATE[1]: 0.5001426719473689 => 0.5001617584303107
STATE[0]: 0.5001040306179318 => 0.5001098033991698
Explore, not updating state value function
Explore, not updating state value function
Explore, not updating state value function
STATE[5]: 0.4955 => 0.49595
STATE[4]: 0.4991 => 0.498785
STATE[3]: 0.49930963784661164 => 0.49925717406195047
STATE[2]: 0.49971193979668704 => 0.4996664632232134
STATE[1]: 0.49984910061655136 => 0.4998308368772176
STATE[0]: 0.49987985041744176 => 0.49987494906341934
STATE[6]: 0.5 => 0.55
STATE[5]: 0.50405 => 0.508645
STATE[4]: 0.501215 => 0.501958
STATE[3]: 0.5007428259380495 => 0.5008643433442446
STATE[2]: 0.5003335367767865 => 0.5003866174335323
STATE[1]: 0.5001617584303107 => 0.5001842443306329
STATE[0]: 0.5001098033991698 => 0.5001172474923161
STATE[6]: 0.5 => 0.45
STATE[5]: 0.49595 => 0.491355
STATE[4]: 0.498785 => 0.498042
STATE[3]: 0.49925717406195047 => 0.49913565665575543
STATE[2]: 0.4996664632232134 => 0.4996133825664676
STATE[1]: 0.4998308368772176 => 0.4998090914461426
STATE[0]: 0.49987494906341934 => 0.49986836330169165
STATE[6]: 0.5 => 0.55
STATE[5]: 0.5095000000000001 => 0.5135500000000001
STATE[4]: 0.5013667900000001 => 0.502585111
STATE[3]: 0.5008643433442446 => 0.5010364201098201
STATE[2]: 0.5003866174335323 => 0.5004515977011611
STATE[1]: 0.5001842443306329 => 0.5002109796676857
STATE[0]: 0.5001172474923161 => 0.500126620709853
STATE[6]: 0.5 => 0.45
STATE[5]: 0.4905 => 0.48645
STATE[4]: 0.49863321000000005 => 0.497414889
STATE[3]: 0.49913565665575543 => 0.4989635798901799
STATE[2]: 0.4996133825664676 => 0.49954840229883885
STATE[1]: 0.4998090914461426 => 0.4997830225314122
```

Explore, not updating state value function
Explore, not updating state value function
Explore, not updating state value function
Explore, not updating state value function
Explore, not updating state value function

`...> 0.4998598292246637`
`...ue function`
`...ue function`

```
STATE[3]: 0.505 => 0.5045
STATE[2]: 0.5004145 => 0.5008230499999999
STATE[1]: 0.5002109796676857 => 0.5002721867009171
STATE[0]: 0.500126620709853 => 0.5001411773089595
Explore, not updating state value function
Explore, not updating state value function
Explore, not updating state value function
Explore, not updating state value function
Explore, not updating state value function
STATE[3]: 0.495 => 0.4955
STATE[2]: 0.49958549999999996 => 0.49917694999999995
STATE[1]: 0.4997830225314122 => 0.499722415278271
STATE[0]: 0.4998598292246637 => 0.49984608783002443
STATE[6]: 0.5 => 0.55
STATE[5]: 0.5095000000000001 => 0.5135500000000001
STATE[4]: 0.5013727860249999 => 0.5025905074224999
STATE[3]: 0.5010364201098201 => 0.501191828841088
STATE[2]: 0.5004515977011611 => 0.5005256208151538
STATE[1]: 0.5002721867009171 => 0.5002975301123408
STATE[0]: 0.5001411773089595 => 0.5001568125892976
STATE[6]: 0.5 => 0.45
STATE[5]: 0.4905 => 0.48645
STATE[4]: 0.498627213975 => 0.4974094925775
STATE[3]: 0.4989635798901799 => 0.4988081711589119
STATE[2]: 0.49954840229883885 => 0.49947437918484616
STATE[1]: 0.499722415278271 => 0.4996976116689285
STATE[0]: 0.49984608783002443 => 0.49983124021391484
STATE[4]: 0.5 => 0.55
STATE[3]: 0.5 => 0.505
STATE[2]: 0.499955 => 0.5004595
STATE[1]: 0.5002975301123408 => 0.5003137271011067
STATE[0]: 0.5001568125892976 => 0.5001725040404785
STATE[4]: 0.5 => 0.45
STATE[3]: 0.5 => 0.495
STATE[2]: 0.499545 => 0.4990905
STATE[1]: 0.4996976116689285 => 0.4996369005020357
STATE[0]: 0.49983124021391484 => 0.4998118062427269
STATE[4]: 0.5 => 0.55
STATE[3]: 0.505 => 0.5095000000000001
STATE[2]: 0.500455 => 0.5013595
STATE[1]: 0.5003137271011067 => 0.500418304390996
STATE[0]: 0.5001725040404785 => 0.5001970840755303
STATE[4]: 0.5 => 0.45
STATE[3]: 0.495 => 0.4905
STATE[2]: 0.499545 => 0.49864050000000004
STATE[1]: 0.4996369005020357 => 0.4995372604518321
STATE[0]: 0.4998118062427269 => 0.49978435166363744
STATE[4]: 0.5 => 0.55
STATE[3]: 0.505 => 0.5095000000000001
STATE[2]: 0.50049955 => 0.501399595
STATE[1]: 0.500418304390996 => 0.5005164334518963
STATE[0]: 0.5001970840755303 => 0.500229019013167
```

Saved successfully!   ✕

```
                                       040499999997
STATE[1]: 0.4995372604518321 => 0.4994435749066489
STATE[0]: 0.49978435166363744 => 0.4997502739879386
STATE[5]: 0.5 => 0.45
```

```
STATE[4]: 0.5 => 0.495
STATE[3]: 0.505 => 0.504
STATE[2]: 0.5005 => 0.5008499999999999
STATE[1]: 0.5005164334518963 => 0.5005497901067066
STATE[0]: 0.500229019013167 => 0.5002610961225209
STATE[5]: 0.5 => 0.55
STATE[4]: 0.5 => 0.505
STATE[3]: 0.495 => 0.496
STATE[2]: 0.4995 => 0.49915
STATE[1]: 0.4994435749066489 => 0.499414217415984
STATE[0]: 0.4997502739879386 => 0.49971666833074313
STATE[4]: 0.5 => 0.55
STATE[3]: 0.505 => 0.5095000000000001
STATE[2]: 0.5005 => 0.5014
STATE[1]: 0.5005497901067066 => 0.5006348110960359
STATE[0]: 0.5002610961225209 => 0.5002984676198724
STATE[4]: 0.5 => 0.45
STATE[3]: 0.495 => 0.4905
STATE[2]: 0.4995 => 0.4986
STATE[1]: 0.499414217415984 => 0.4993327956743856
STATE[0]: 0.49971666833074313 => 0.49967828106510737
STATE[4]: 0.5 => 0.55
STATE[3]: 0.505 => 0.5095000000000001
STATE[2]: 0.5005 => 0.5014
STATE[1]: 0.5006348110960359 => 0.5007113299864323
STATE[0]: 0.5002984676198724 => 0.5003397538565284
STATE[4]: 0.5 => 0.45
STATE[3]: 0.495 => 0.4905
STATE[2]: 0.4995 => 0.4986
STATE[1]: 0.4993327956743856 => 0.499259516106947
STATE[0]: 0.49967828106510737 => 0.4996364045692913
STATE[6]: 0.5 => 0.55
STATE[5]: 0.5095000000000001 => 0.5135500000000001
STATE[4]: 0.5014 => 0.5026149999999999
STATE[3]: 0.501191828841088 => 0.5013341459569792
STATE[2]: 0.5005256208151538 => 0.5006064733293364
STATE[1]: 0.5007113299864323 => 0.5007008443207227
STATE[0]: 0.5003397538565284 => 0.5003758629029478
STATE[6]: 0.5 => 0.45
STATE[5]: 0.4905 => 0.48645
STATE[4]: 0.4986 => 0.49738499999999997
STATE[3]: 0.4988081711589119 => 0.4986658540430207
STATE[2]: 0.49947437918484616 => 0.4993935266706636
STATE[1]: 0.499259516106947 => 0.4992729171633187
STATE[0]: 0.4996364045692913 => 0.49960005582869405
STATE[8]: 0.5 => 0.55
STATE[7]: 0.5 => 0.505
STATE[6]: 0.5 => 0.5005
STATE[5]: 0.5095000000000001 => 0.5086
STATE[4]: 0.5014 => 0.50212
STATE[3]: 0.5013341459569792 => 0.5014127313612813
STATE[2]: 0.5006064733293364 => 0.5006870991325308
                                  => 0.5006994698019035
                                  => 0.5004082235928433
STATE[8]: 0.5 => 0.45
STATE[7]: 0.5 => 0.495
STATE[6]: 0.5 => 0.4995
STATE[5]: 0.4905 => 0.4914
```

Saved successfully! ✕

```
STATE[4]: 0.4986 => 0.49788
STATE[3]: 0.4986658540430207 => 0.49858726863871866
STATE[2]: 0.4993935266706636 => 0.4993129008674691
STATE[1]: 0.4992729171633187 => 0.4992769155337337
STATE[0]: 0.49960005582869405 => 0.499567741799198
STATE[6]: 0.5 => 0.55
STATE[5]: 0.5095000000000001 => 0.5135500000000001
STATE[4]: 0.5014 => 0.5026149999999999
STATE[3]: 0.5014127313612813 => 0.5015329582251531
STATE[2]: 0.5006870991325308 => 0.500771685041793
STATE[1]: 0.5006994698019035 => 0.5007066913258924
STATE[0]: 0.5004082235928433 => 0.5004380703661482
STATE[6]: 0.5 => 0.45
STATE[5]: 0.4905 => 0.48645
STATE[4]: 0.4986 => 0.49738499999999997
STATE[3]: 0.49858726863871866 => 0.4984670417748468
STATE[2]: 0.4993129008674691 => 0.4992831495820685
STATE[1]: 0.4992769155337337 => 0.49927205547618103
STATE[0]: 0.499567741799198 => 0.4995381731668963
Agent is taking an explore action
Explore, not updating state value function
Explore, not updating state value function
Explore, not updating state value function
STATE[5]: 0.508645 => 0.5077805
STATE[4]: 0.501958 => 0.50254025
STATE[3]: 0.5015329582251531 => 0.5016336874026378
STATE[2]: 0.500771685041793 => 0.5008578852778776
STATE[1]: 0.5007066913258924 => 0.5007218107210909
STATE[0]: 0.5004380703661482 => 0.5004664444016425
Explore, not updating state value function
Explore, not updating state value function
Explore, not updating state value function
STATE[5]: 0.491355 => 0.4922195
STATE[4]: 0.498042 => 0.49745975
STATE[3]: 0.4984670417748468 => 0.49836631259736214
STATE[2]: 0.4992831495820685 => 0.4991421147221224
STATE[1]: 0.49927205547618103 => 0.49925906140077514
STATE[0]: 0.4995381731668963 => 0.4995102619902842
STATE[4]: 0.5 => 0.55
STATE[3]: 0.5045 => 0.50905
STATE[2]: 0.5008230499999999 => 0.5016457449999999
STATE[1]: 0.5007218107210909 => 0.5008142041489818
STATE[0]: 0.5004664444016425 => 0.5005012203763763
STATE[4]: 0.5 => 0.45
STATE[3]: 0.4955 => 0.49095
STATE[2]: 0.49917694999999995 => 0.49835425499999997
STATE[1]: 0.49925906140077514 => 0.49916858076069764
STATE[0]: 0.4995102619902842 => 0.49947609386732555
STATE[5]: 0.5 => 0.45
STATE[4]: 0.495 => 0.4905
STATE[3]: 0.504 => 0.50265
STATE[                      ]     0.5010299999999999
STATE                           0.5008357837340837
STATE                           0.500534676712147
STATE[5]: 0.5 => 0.55
STATE[4]: 0.505 => 0.5095000000000001
STATE[3]: 0.496 => 0.49735
```

Saved successfully!   ×

```
STATE[2]: 0.49915 => 0.4989699999999997
STATE[1]: 0.49916858076069764 => 0.4991487226846279
STATE[0]: 0.49947609386732555 => 0.49944335674905577
Explore, not updating state value function
STATE[7]: 0.505 => 0.5045
STATE[6]: 0.5005 => 0.5008999999999999
STATE[5]: 0.5086 => 0.50783
STATE[4]: 0.50212 => 0.502691
STATE[3]: 0.5016336874026378 => 0.501739418662374
STATE[2]: 0.5008578852778776 => 0.5009460386163272
STATE[1]: 0.5008357837340837 => 0.5008468092223081
STATE[0]: 0.500534676712147 => 0.5005658899631631
Explore, not updating state value function
STATE[7]: 0.495 => 0.4955
STATE[6]: 0.4995 => 0.4991
STATE[5]: 0.4914 => 0.49217
STATE[4]: 0.49788 => 0.497309
STATE[3]: 0.49836631259736214 => 0.4982605813376259
STATE[2]: 0.4991421147221224 => 0.49905396138367275
STATE[1]: 0.4991487226846279 => 0.4991392465545324
STATE[0]: 0.49944335674905577 => 0.49941294572960343
STATE[4]: 0.5 => 0.55
STATE[3]: 0.505 => 0.5095000000000001
STATE[2]: 0.5004595 => 0.50136355
STATE[1]: 0.5008468092223081 => 0.5008984833000772
STATE[0]: 0.5005658899631631 => 0.5005991492968545
STATE[4]: 0.5 => 0.45
STATE[3]: 0.495 => 0.4905
STATE[2]: 0.4990905 => 0.49823145
STATE[1]: 0.4991392465545324 => 0.49904846689907917
STATE[0]: 0.49941294572960343 => 0.499376497846551
STATE[8]: 0.5 => 0.55
STATE[7]: 0.5 => 0.505
STATE[6]: 0.5 => 0.5005
STATE[5]: 0.5077805 => 0.50705245
STATE[4]: 0.50254025 => 0.50299147
STATE[3]: 0.501739418662374 => 0.5018646237961366
STATE[2]: 0.5009460386163272 => 0.5010378971343081
STATE[1]: 0.5008984833000772 => 0.5009124246835004
STATE[0]: 0.5005991492968545 => 0.5006304768355191
STATE[8]: 0.5 => 0.45
STATE[7]: 0.5 => 0.495
STATE[6]: 0.5 => 0.4995
STATE[5]: 0.4922195 => 0.49294754999999996
STATE[4]: 0.49745975 => 0.49700853
STATE[3]: 0.4982605813376259 => 0.49813537620386333
STATE[2]: 0.49905396138367275 => 0.4989621028656918
STATE[1]: 0.49904846689907917 => 0.4990398304957404
STATE[0]: 0.499376497846551 => 0.49934283111147
STATE[5]: 0.5 => 0.45
STATE[4]: 0.4905 => 0.48645
STATE[3]: 0.50265 => 0.5010300000000001
                                         0.5010299999999999
                                         0.5009241822151503
                                         0.5006598473734822
STATE[5]: 0.5 => 0.55
STATE[4]: 0.5095000000000001 => 0.5135500000000001
STATE[3]: 0.49735 => 0.49897
```

Saved successfully! ✕

```
STATE[2]: 0.4989699999999997 => 0.4989699999999997
STATE[1]: 0.4990398304957404 => 0.4990328474461664
STATE[0]: 0.49934283111147 => 0.4993118327449396
STATE[7]: 0.5 => 0.45
STATE[6]: 0.5 => 0.495
STATE[5]: 0.5 => 0.4995
STATE[4]: 0.48645 => 0.487755
STATE[3]: 0.5010300000000001 => 0.4997025000000005
STATE[2]: 0.5010299999999999 => 0.5008972499999998
STATE[1]: 0.5009241822151503 => 0.5009214889936353
STATE[0]: 0.5006598473734822 => 0.5006860115354975
STATE[7]: 0.5 => 0.55
STATE[6]: 0.5 => 0.505
STATE[5]: 0.5 => 0.5005
STATE[4]: 0.5135500000000001 => 0.5122450000000001
STATE[3]: 0.49897 => 0.5002975000000001
STATE[2]: 0.4989699999999997 => 0.49910275
STATE[1]: 0.4990328474461664 => 0.49903983770154975
STATE[0]: 0.4993118327449396 => 0.49928463324060063
STATE[4]: 0.5 => 0.55
STATE[3]: 0.5 => 0.505
STATE[2]: 0.5008972499999998 => 0.5013075249999999
STATE[1]: 0.5009214889936353 => 0.5009600925942718
STATE[0]: 0.5006860115354975 => 0.500713419641375
STATE[4]: 0.5 => 0.45
STATE[3]: 0.5 => 0.495
STATE[2]: 0.49910275 => 0.49869247499999997
STATE[1]: 0.49903983770154975 => 0.49900510143139476
STATE[0]: 0.49928463324060063 => 0.49925668005968005
STATE[6]: 0.55 => 0.5950000000000001
STATE[5]: 0.5135500000000001 => 0.521695
STATE[4]: 0.502585111 => 0.5044960999
STATE[3]: 0.5018646237961366 => 0.502127771406523
STATE[2]: 0.5010378971343081 => 0.5011468845615297
STATE[1]: 0.5009600925942718 => 0.5009787717909976
STATE[0]: 0.500713419641375 => 0.5007399548563372
STATE[6]: 0.45 => 0.405
STATE[5]: 0.48645 => 0.478305
STATE[4]: 0.497414889 => 0.4955039001
STATE[3]: 0.49813537620386333 => 0.497872228593477
STATE[2]: 0.4989621028656918 => 0.49885311543847033
STATE[1]: 0.49900510143139476 => 0.4989899028321023
STATE[0]: 0.49925668005968005 => 0.4992300023369223
Agent is taking an explore action
STATE[4]: 0.5 => 0.55
STATE[3]: 0.5 => 0.505
Explore, not updating state value function
STATE[1]: 0.5009787717909976 => 0.5009955830680508
STATE[0]: 0.5007399548563372 => 0.5007655176775085
STATE[4]: 0.5 => 0.45
STATE[3]: 0.5 => 0.495
STATE[2]: 0.49885311543847033 => 0.4984678038946233
                                 0.49893769293835444
                                 0.4992007713970655
Agent is taking an explore action
STATE[4]: 0.5 => 0.55
STATE[3]: 0.505 => 0.5095000000000001
STATE[2]: 0.5013075249999999 => 0.5021267724999999
```

```
STATE[2]: 0.501507924999999 => 0.502126772499999
STATE[1]: 0.5009955830680508 => 0.5011087020112457
STATE[0]: 0.5007655176775085 => 0.5007998361108823
STATE[4]: 0.5 => 0.45
Explore, not updating state value function
STATE[2]: 0.49869247499999997 => 0.49832322749999997
STATE[1]: 0.49893769293835444 => 0.498876246394519
STATE[0]: 0.4992007713970655 => 0.49916831889681085
Agent is taking an explore action
STATE[4]: 0.5 => 0.55
STATE[3]: 0.5095000000000001 => 0.5135500000000001
STATE[2]: 0.5013595 => 0.50257855
STATE[1]: 0.5011087020112457 => 0.5012556868101211
Explore, not updating state value function
STATE[4]: 0.5 => 0.45
STATE[3]: 0.4905 => 0.48645
STATE[2]: 0.49864050000000004 => 0.49742145000000004
STATE[1]: 0.498876246394519 => 0.4987307667550671
STATE[0]: 0.49916831889681085 => 0.49912456368263647
STATE[4]: 0.5 => 0.55
STATE[3]: 0.5095000000000001 => 0.5135500000000001
STATE[2]: 0.501399595 => 0.5026146355
STATE[1]: 0.5012556868101211 => 0.501391581679109
STATE[0]: 0.5007998361108823 => 0.500859010667705
STATE[4]: 0.5 => 0.45
STATE[3]: 0.4905 => 0.48645
STATE[2]: 0.49860040499999997 => 0.4973853645
STATE[1]: 0.4987307667550671 => 0.49859622652956037
STATE[0]: 0.49912456368263647 => 0.49907172996732885
STATE[4]: 0.5 => 0.55
STATE[3]: 0.5095000000000001 => 0.5135500000000001
STATE[2]: 0.5014 => 0.5026149999999999
STATE[1]: 0.501391581679109 => 0.5015139235111981
STATE[0]: 0.500859010667705 => 0.5009245019520543
STATE[4]: 0.5 => 0.45
STATE[3]: 0.4905 => 0.48645
STATE[2]: 0.4986 => 0.49738499999999997
STATE[1]: 0.49859622652956037 => 0.4984751038766043
STATE[0]: 0.49907172996732885 => 0.4990120673582564
STATE[4]: 0.5 => 0.55
STATE[3]: 0.5095000000000001 => 0.5135500000000001
STATE[2]: 0.5014 => 0.5026149999999999
STATE[1]: 0.5015139235111981 => 0.5016240311600783
STATE[0]: 0.5009245019520543 => 0.5009944548728567
STATE[4]: 0.5 => 0.45
STATE[3]: 0.4905 => 0.48645
STATE[2]: 0.4986 => 0.49738499999999997
STATE[1]: 0.4984751038766043 => 0.4983660934889439
STATE[0]: 0.4990120673582564 => 0.4989474699713251
STATE[4]: 0.5 => 0.55
STATE[3]: 0.505 => 0.5095000000000001
STATE[2]: 0.5011468845615297 => 0.5019821961053768
                                => 0.5016598476546081
                          ×     => 0.5010609941510318

STATE[3]: 0.495 => 0.4905
STATE[2]: 0.4984678038946233 => 0.49767102350516096
STATE[1]: 0.4983660934889439 => 0.49829658649056563
```

```
STATE[0]: 0.4989474699713251 => 0.49888238162324916
STATE[6]: 0.5 => 0.55
STATE[5]: 0.5 => 0.505
STATE[4]: 0.5 => 0.5005
STATE[3]: 0.50905 => 0.508195
STATE[2]: 0.5016457449999999 => 0.5023006704999999
STATE[1]: 0.5016598476546081 => 0.5017239299391473
STATE[0]: 0.5010609941510318 => 0.5011272877298434
STATE[6]: 0.5 => 0.45
STATE[5]: 0.5 => 0.495
STATE[4]: 0.5 => 0.4995
STATE[3]: 0.49095 => 0.491805
STATE[2]: 0.49835425499999997 => 0.49769932949999995
STATE[1]: 0.49829658649056563 => 0.4982368607915091
STATE[0]: 0.49888238162324916 => 0.49881782954007514
STATE[4]: 0.5 => 0.55
STATE[3]: 0.5095000000000001 => 0.5135500000000001
STATE[2]: 0.5021267724999999 => 0.50326909525
STATE[1]: 0.5017239299391473 => 0.5018784464702325
STATE[0]: 0.5011272877298434 => 0.5012024036038822
STATE[4]: 0.5 => 0.45
STATE[3]: 0.495 => 0.4905
STATE[2]: 0.49832322749999997 => 0.49754090474999996
STATE[1]: 0.4982368607915091 => 0.49816726518735815
STATE[0]: 0.49881782954007514 => 0.4987527731048034
STATE[6]: 0.5 => 0.55
STATE[5]: 0.5 => 0.505
STATE[4]: 0.5 => 0.5005
STATE[3]: 0.5095000000000001 => 0.5086
STATE[2]: 0.50136355 => 0.502087195
STATE[1]: 0.5018784464702325 => 0.5018993213232092
STATE[0]: 0.5012024036038822 => 0.501272095375815
STATE[6]: 0.5 => 0.45
STATE[5]: 0.5 => 0.495
STATE[4]: 0.5 => 0.4995
STATE[3]: 0.4905 => 0.4914
STATE[2]: 0.49823145 => 0.497548305
STATE[1]: 0.49816726518735815 => 0.49810536916862236
STATE[0]: 0.4987527731048034 => 0.49868803271118534
STATE[4]: 0.5 => 0.55
STATE[3]: 0.508195 => 0.5123755
STATE[2]: 0.5023006704999999 => 0.5033081534499999
STATE[1]: 0.5018993213232092 => 0.5020402045358883
STATE[0]: 0.501272095375815 => 0.5013489062918223
STATE[4]: 0.5 => 0.45
STATE[3]: 0.491805 => 0.4876245
STATE[2]: 0.49769932949999995 => 0.49669184655
STATE[1]: 0.49810536916862236 => 0.4979640169067601
STATE[0]: 0.49868803271118534 => 0.4986156311307428
Agent is taking an explore action
STATE[7]: 0.5 => 0.45
Explore, not updating state value function
```

Saved successfully! ✕

```
                                          ue function
                                          ue function
                            => 0.5085500000000001
STATE[2]: 0.5019821961053768 => 0.5026389764948391
STATE[1]: 0.5020402045358883 => 0.5021000817317833
STATE[0]: 0.5013489062918223 => 0.5014240238358184
```

```
STATE[7]: 0.5 => 0.55
STATE[6]: 0.5 => 0.505
STATE[5]: 0.5 => 0.5005
STATE[4]: 0.5 => 0.50005
STATE[3]: 0.4905 => 0.491455
STATE[2]: 0.49767102350516096 => 0.49704942115464484
STATE[1]: 0.4979640169067601 => 0.4978725573315486
STATE[0]: 0.4986156311307428 => 0.4985413237508234
STATE[4]: 0.5 => 0.55
STATE[3]: 0.5086 => 0.5127400000000001
STATE[2]: 0.502087195 => 0.5031524755
STATE[1]: 0.5021000817317833 => 0.502205321108605
STATE[0]: 0.5014240238358184 => 0.501502153563097
STATE[4]: 0.5 => 0.45
STATE[3]: 0.4914 => 0.48726
STATE[2]: 0.497548305 => 0.4965194745
STATE[1]: 0.4978725573315486 => 0.4977372490483938
STATE[0]: 0.4985413237508234 => 0.4984609162805804
STATE[4]: 0.5 => 0.55
STATE[3]: 0.5135500000000001 => 0.5171950000000001
STATE[2]: 0.50326909525 => 0.5046616857249999
STATE[1]: 0.502205321108605 => 0.5024509575702445
STATE[0]: 0.501502153563097 => 0.5015970339638117
STATE[4]: 0.5 => 0.45
STATE[3]: 0.4905 => 0.48645
STATE[2]: 0.49754090474999996 => 0.49643181427499994
STATE[1]: 0.4977372490483938 => 0.4976067055710544
STATE[0]: 0.4984609162805804 => 0.4983754952096278
STATE[5]: 0.5 => 0.45
STATE[4]: 0.5 => 0.495
STATE[3]: 0.5135500000000001 => 0.511695
STATE[2]: 0.50257855 => 0.5034901949999999
STATE[1]: 0.5024509575702445 => 0.50255488131322
STATE[0]: 0.5015970339638117 => 0.5016928186987526
STATE[5]: 0.5 => 0.55
STATE[4]: 0.5 => 0.505
STATE[3]: 0.48645 => 0.488305
STATE[2]: 0.49742145000000004 => 0.496509805
STATE[1]: 0.4976067055710544 => 0.497497015513949
STATE[0]: 0.4983754952096278 => 0.4982876472400599
STATE[4]: 0.5 => 0.55
STATE[3]: 0.5135500000000001 => 0.5171950000000001
STATE[2]: 0.5026146355 => 0.5040726719499999
STATE[1]: 0.50255488131322 => 0.502706660376898
STATE[0]: 0.5016928186987526 => 0.5017942028665672
STATE[4]: 0.5 => 0.45
STATE[3]: 0.48645 => 0.482805
STATE[2]: 0.4973853645 => 0.49592732805
STATE[1]: 0.497497015513949 => 0.49734004676755406
STATE[0]: 0.4982876472400599 => 0.49819288719280935
STATE[4]: 0.5 => 0.55
STATE[3]: 0.5135500000000001 => 0.5171950000000001
                              0.504073
                              0.5028432943392083
STATE[0]: 0.5017942028665672 => 0.5018991120138313
STATE[4]: 0.5 => 0.45
STATE[3]: 0.48645 => 0.482805
STATE[2]: 0.4973849999999997 => 0.4959269999999995
```

Saved successfully!   ×

```
STATE[2]: 0.49738499999997 => 0.49552899999999
STATE[1]: 0.49734004676755406 => 0.4971987420907986
STATE[0]: 0.49819288719280935 => 0.4980934726826083
Explore, not updating state value function
STATE[7]: 0.505 => 0.5045
STATE[6]: 0.5 => 0.50045
STATE[5]: 0.5 => 0.500045
STATE[4]: 0.5 => 0.5000045
STATE[3]: 0.5135500000000001 => 0.5121954500000001
STATE[2]: 0.5026149999999999 => 0.503573045
STATE[1]: 0.5028432943392083 => 0.50291626940852874
STATE[0]: 0.5018991120138313 => 0.5020008277529769
Explore, not updating state value function
STATE[7]: 0.495 => 0.4955
STATE[6]: 0.5 => 0.49955
STATE[5]: 0.5 => 0.499955
STATE[4]: 0.5 => 0.4999955
STATE[3]: 0.48645 => 0.48780455
STATE[2]: 0.49738499999999997 => 0.49642695499999995
STATE[1]: 0.4971987420907986 => 0.49712156338171876
STATE[0]: 0.4980934726826083 => 0.4979962817525193
STATE[6]: 0.5 => 0.55
STATE[5]: 0.5 => 0.505
STATE[4]: 0.5 => 0.5005
STATE[3]: 0.5085500000000001 => 0.507745
STATE[2]: 0.5026389764948391 => 0.5031495788453553
STATE[1]: 0.5029162694052874 => 0.5029396003492942
STATE[0]: 0.5020008277529769 => 0.5020947050126087
STATE[6]: 0.5 => 0.45
STATE[5]: 0.5 => 0.495
STATE[4]: 0.50005 => 0.499545
STATE[3]: 0.491455 => 0.492264
STATE[2]: 0.49704942115464484 => 0.49657087903918035
STATE[1]: 0.49712156338171876 => 0.4970664949474649
STATE[0]: 0.4979962817525193 => 0.49790330307201386
STATE[7]: 0.5 => 0.45
STATE[6]: 0.5 => 0.495
STATE[5]: 0.505 => 0.504
STATE[4]: 0.5005 => 0.5008499999999999
STATE[3]: 0.5123755 => 0.51122295
STATE[2]: 0.5033081534499999 => 0.504099633105
STATE[1]: 0.5029396003492942 => 0.5030556036248648
STATE[0]: 0.5020947050126087 => 0.5021907948738342
STATE[7]: 0.5 => 0.55
STATE[6]: 0.5 => 0.505
STATE[5]: 0.495 => 0.496
STATE[4]: 0.4995 => 0.49915
STATE[3]: 0.4876245 => 0.48877705
STATE[2]: 0.49669184655 => 0.495900366895
STATE[1]: 0.4970664949474649 => 0.4969498821422184
STATE[0]: 0.49790330307201386 => 0.4978079609790343
STATE[4]: 0.5 => 0.55
```

Saved successfully!   ✕   0.5040316709608197
                                    0.5031532103584603

```
STATE[0]: 0.5021907948738342 => 0.5022870364222969
STATE[4]: 0.5 => 0.45
STATE[3]: 0.492264 => 0.48803759999999996
```

```
STATE[2]: 0.49657087903918035 => 0.4957175511352623
STATE[1]: 0.4969498821422184 => 0.4968266490415228
STATE[0]: 0.4978079609790343 => 0.49770982978528316
STATE[4]: 0.5 => 0.55
STATE[3]: 0.5127400000000001 => 0.5164660000000001
STATE[2]: 0.5031524755 => 0.50448382795
STATE[1]: 0.5031532103584603 => 0.5032862721176142
STATE[0]: 0.5022870364222969 => 0.5023869599918286
STATE[4]: 0.5 => 0.45
STATE[3]: 0.48726 => 0.483534
STATE[2]: 0.4965194745 => 0.49522092705
STATE[1]: 0.4968266490415228 => 0.4966660768423705
STATE[0]: 0.49770982978528316 => 0.4976054544909919
STATE[8]: 0.5 => 0.55
STATE[7]: 0.5 => 0.505
STATE[6]: 0.5 => 0.5005
STATE[5]: 0.5 => 0.50005
STATE[4]: 0.495 => 0.495505
STATE[3]: 0.511695 => 0.510076
STATE[2]: 0.5034901949999999 => 0.5041487754999999
STATE[1]: 0.5032862721176142 => 0.5033725224558528
STATE[0]: 0.5023869599918286 => 0.502485516238231
STATE[8]: 0.5 => 0.45
STATE[7]: 0.5 => 0.495
STATE[6]: 0.5 => 0.4995
STATE[5]: 0.5 => 0.49995
STATE[4]: 0.505 => 0.504495
STATE[3]: 0.488305 => 0.48992399999999997
STATE[2]: 0.496509805 => 0.4958512245
STATE[1]: 0.4966660768423705 => 0.49658459160813345
STATE[0]: 0.4976054544909919 => 0.4975033682027061
STATE[6]: 0.5 => 0.55
STATE[5]: 0.5 => 0.505
STATE[4]: 0.5 => 0.5005
STATE[3]: 0.5171950000000001 => 0.5155255000000001
STATE[2]: 0.5046616857249999 => 0.5057480671525
STATE[1]: 0.5033725224558528 => 0.5036100769255175
STATE[0]: 0.502485516238231 => 0.5025979723069597
STATE[6]: 0.5 => 0.45
STATE[5]: 0.5 => 0.495
STATE[4]: 0.5 => 0.4995
STATE[3]: 0.48645 => 0.487755
STATE[2]: 0.49643181427499994 => 0.49556413284749995
STATE[1]: 0.49658459160813345 => 0.4964825457320701
STATE[0]: 0.4975033682027061 => 0.4974012859556425
STATE[4]: 0.5 => 0.55
STATE[3]: 0.5121954500000001 => 0.5159759050000001
STATE[2]: 0.503573045 => 0.504813331
STATE[1]: 0.5036100769255175 => 0.5037304023329657
STATE[0]: 0.5025979723069597 => 0.5027112153095603
STATE[4]: 0.5 => 0.45
STATE[3]: 0.48780455 => 0.484024095
             => 0.49518666899999997
             0.4963529580588631
             0.49729645316596455
STATE[6]: 0.5 => 0.55
STATE[5]: 0.5 => 0.505
STATE[4]: 0.5 => 0.5005
```

Saved successfully!   ✕

```
STATE[3]: 0.5171950000000001 => 0.5155255000000001
STATE[2]: 0.5040726719499999 => 0.505217954755
STATE[1]: 0.5037304023329657 => 0.5038791575751691
STATE[0]: 0.5027112153095603 => 0.5028280095361212
STATE[6]: 0.5 => 0.45
STATE[5]: 0.5 => 0.495
STATE[4]: 0.5 => 0.4995
STATE[3]: 0.482805 => 0.4844745
STATE[2]: 0.49592732805 => 0.494782045245
STATE[1]: 0.4963529580588631 => 0.4961958667774768
STATE[0]: 0.49729645316596455 => 0.49718639452711577
STATE[5]: 0.5 => 0.45
STATE[4]: 0.5 => 0.495
STATE[3]: 0.5171950000000001 => 0.5149755
STATE[2]: 0.504073 => 0.50516325
STATE[1]: 0.5038791575751691 => 0.5040075668176522
STATE[0]: 0.5028280095361212 => 0.5029459652642743
STATE[5]: 0.5 => 0.55
STATE[4]: 0.5 => 0.505
STATE[3]: 0.482805 => 0.48502449999999997
STATE[2]: 0.49592699999999995 => 0.49483674999999994
STATE[1]: 0.4961958667774768 => 0.4960599550997291
STATE[0]: 0.49718639452711577 => 0.4970737505843771
STATE[7]: 0.5 => 0.45
STATE[6]: 0.495 => 0.4905
STATE[5]: 0.504 => 0.50265
STATE[4]: 0.5008499999999999 => 0.5010299999999999
STATE[3]: 0.51122295 => 0.510203655
STATE[2]: 0.504099633105 => 0.5047100352945
STATE[1]: 0.5040075668176522 => 0.504077813665337
STATE[0]: 0.5029459652642743 => 0.5030591501043805
STATE[7]: 0.5 => 0.55
STATE[6]: 0.505 => 0.5095000000000001
STATE[5]: 0.496 => 0.49735
STATE[4]: 0.49915 => 0.49896999999999997
STATE[3]: 0.48877705 => 0.48979634499999997
STATE[2]: 0.495900366895 => 0.4952899647055
STATE[1]: 0.4960599550997291 => 0.4959829560603062
STATE[0]: 0.4970737505843771 => 0.49696467113197
STATE[7]: 0.5 => 0.45
STATE[6]: 0.5 => 0.495
STATE[5]: 0.50005 => 0.499545
STATE[4]: 0.495505 => 0.495909
STATE[3]: 0.510076 => 0.5086592999999999
STATE[2]: 0.5041487754999999 => 0.5045998279499999
STATE[1]: 0.504077813665337 => 0.5041300150938033
STATE[0]: 0.5030591501043805 => 0.5031662366033228
STATE[7]: 0.5 => 0.55
STATE[6]: 0.5 => 0.505
STATE[5]: 0.49995 => 0.500455
STATE[4]: 0.504495 => 0.5040910000000001
STATE[3]: 0.4890380000000007 => 0.49134069999999996
                          40017205000003
                          0.49592467765927556
STATE[0]: 0.49696467113197 => 0.49686067178470056
STATE[4]: 0.5 => 0.55
STATE[3]: 0.5119705 => 0.51577345
STATE[2]: 0.504031670068197 => 0.5052058488647377
```

Saved successfully!   ×

```
STATE[2]: 0.5040318709608197 -> 0.5052058488647577
STATE[1]: 0.5041300150938033 => 0.5042375984708968
STATE[0]: 0.5031662366033228 => 0.5032733727900802
STATE[4]: 0.5 => 0.45
STATE[3]: 0.48803759999999996 => 0.48423383999999997
STATE[2]: 0.4957175511352623 => 0.49456918002173605
STATE[1]: 0.49592467765927556 => 0.4957891278955216
STATE[0]: 0.49686067178470056 => 0.49675351739578266
STATE[6]: 0.5 => 0.55
STATE[5]: 0.505 => 0.5095000000000001
STATE[4]: 0.5005 => 0.5014
STATE[3]: 0.5155255000000001 => 0.5141129500000001
STATE[2]: 0.5057480671525 => 0.50658455543725
STATE[1]: 0.5042375984708968 => 0.5044722941675321
STATE[0]: 0.5032733727900802 => 0.5033932649278254
STATE[6]: 0.5 => 0.45
STATE[5]: 0.495 => 0.4905
STATE[4]: 0.4995 => 0.4986
STATE[3]: 0.487755 => 0.4888395
STATE[2]: 0.49556413284749995 => 0.49489166956274994
STATE[1]: 0.4957891278955216 => 0.49569938206224445
STATE[0]: 0.49675351739578266 => 0.49664810386242886
STATE[5]: 0.5 => 0.45
STATE[4]: 0.495909 => 0.4913181
STATE[3]: 0.5086592999999999 => 0.50692518
STATE[2]: 0.5045998279499999 => 0.5048323631549999
STATE[1]: 0.5044722941675321 => 0.504508301066279
STATE[0]: 0.5033932649278254 => 0.5035047685416708
STATE[5]: 0.5 => 0.55
STATE[4]: 0.5040910000000001 => 0.5086819
STATE[3]: 0.49134069999999996 => 0.49307481999999997
STATE[2]: 0.49540017205000003 => 0.49516763684500004
STATE[1]: 0.49569938206224445 => 0.49564620754052
STATE[0]: 0.49664810386242886 => 0.496547914230238
Explore, not updating state value function
Explore, not updating state value function
STATE[6]: 0.4905 => 0.49145
STATE[5]: 0.50265 => 0.50153
STATE[4]: 0.5010299999999999 => 0.5010799999999999
STATE[3]: 0.510203655 => 0.5092912894999999
STATE[2]: 0.5047100352945 => 0.50516816071505
STATE[1]: 0.504508301066279 => 0.504574287031156
STATE[0]: 0.5035047685416708 => 0.5036117203906193
Explore, not updating state value function
Explore, not updating state value function
STATE[6]: 0.5095000000000001 => 0.5085500000000001
STATE[5]: 0.49735 => 0.49847
STATE[4]: 0.49896999999999997 => 0.49892
STATE[3]: 0.48979634499999997 => 0.4907087105
STATE[2]: 0.4952899647055 => 0.49483183928495
STATE[1]: 0.49564620754052 => 0.495564770714963
STATE[0]: 0.496547914230238 => 0.49644959987871046
```
```
                                  0.5198194
                                  6017385155
STATE[1]: 0.504574287031156 => 0.5047185968435405
STATE[0]: 0.5036117203906193 => 0.5037224080359114
STATE[4]: 0.5 => 0.45
```

```
STATE[3]: 0.483534 => 0.4801806
STATE[2]: 0.49522092705 => 0.493716894345
STATE[1]: 0.495564770714963 => 0.4953799830779667
STATE[0]: 0.49644959987871046 => 0.4963426381986361
STATE[8]: 0.5 => 0.55
STATE[7]: 0.5 => 0.505
STATE[6]: 0.5 => 0.5005
STATE[5]: 0.500045 => 0.5000905
STATE[4]: 0.5000045 => 0.5000131
STATE[3]: 0.5159759050000001 => 0.5143796245000001
STATE[2]: 0.504813331 => 0.50576996035
STATE[1]: 0.5047185968435405 => 0.5048237331941864
STATE[0]: 0.5037224080359114 => 0.5038325405517389
STATE[8]: 0.5 => 0.45
STATE[7]: 0.5 => 0.495
STATE[6]: 0.5 => 0.4995
STATE[5]: 0.499955 => 0.4999095
STATE[4]: 0.4999955 => 0.4999869
STATE[3]: 0.484024095 => 0.4856203755
STATE[2]: 0.49518666899999997 => 0.49423003964999995
STATE[1]: 0.4953799830779667 => 0.49526498873517005
STATE[0]: 0.4963426381986361 => 0.4962348732522895
STATE[5]: 0.5 => 0.45
STATE[4]: 0.4913181 => 0.48718628999999997
STATE[3]: 0.50692518 => 0.504951291
STATE[2]: 0.5048323631549999 => 0.5048442559394999
STATE[1]: 0.5048237331941864 => 0.5048257854687177
STATE[0]: 0.5038325405517389 => 0.5039318650434368
STATE[5]: 0.5 => 0.55
STATE[4]: 0.5086819 => 0.51281371
STATE[3]: 0.49307481999999997 => 0.495048709
STATE[2]: 0.49516763684500004 => 0.4951557440605
STATE[1]: 0.49526498873517005 => 0.495254064267703
STATE[0]: 0.4962348732522895 => 0.49613679235383085
STATE[5]: 0.5 => 0.45
STATE[4]: 0.48718628999999997 => 0.483467661
STATE[3]: 0.504951291 => 0.502802928
STATE[2]: 0.5048442559394999 => 0.50464012314555
STATE[1]: 0.5048257854687177 => 0.5048072192364009
STATE[0]: 0.5039318650434368 => 0.5040194004627332
STATE[5]: 0.5 => 0.55
STATE[4]: 0.51281371 => 0.516532339
STATE[3]: 0.495048709 => 0.49719707199999996
STATE[2]: 0.4951557440605 => 0.49535987685445
STATE[1]: 0.495254064267703 => 0.4952646455263777
STATE[0]: 0.49613679235383085 => 0.49604957767108554
Explore, not updating state value function
Explore, not updating state value function
STATE[6]: 0.495 => 0.4955
STATE[5]: 0.499545 => 0.4991405
STATE[4]: 0.483467661 => 0.4850349449
STATE[3]: 0.502802928 => 0.50102612969
                                    .5042787237999949
                                    0.5047543696927603
STATE[0]: 0.5040194004627332 => 0.5040928973857359
Explore, not updating state value function
Explore, not updating state value function
STATE[6]: 0.505 => 0.5045
```

Saved successfully!   ×

```
STATE[5]: 0.500455 => 0.5008595
STATE[4]: 0.516532339 => 0.5149650551
STATE[3]: 0.49719707199999996 => 0.49897387030999996
STATE[2]: 0.49535987685445 => 0.495721276200005
STATE[1]: 0.4952646455263777 => 0.4953103085937405
STATE[0]: 0.49604957767108554 => 0.49597565076335104
STATE[7]: 0.5 => 0.45
STATE[6]: 0.4955 => 0.49095
STATE[5]: 0.4991405 => 0.49832145
STATE[4]: 0.4850349449 => 0.48636359540999996
STATE[3]: 0.50102612969 => 0.49955987626200005
STATE[2]: 0.5042787237999949 => 0.5038068390461955
STATE[1]: 0.5047543696927603 => 0.5046596166281039
STATE[0]: 0.5040928973857359 => 0.5041495693099727
STATE[7]: 0.5 => 0.55
STATE[6]: 0.5045 => 0.50905
STATE[5]: 0.5008595 => 0.5016785500000001
STATE[4]: 0.5149650551 => 0.51363640459
STATE[3]: 0.49897387030999996 => 0.500440123738
STATE[2]: 0.495721276200005 => 0.4961931609538045
STATE[1]: 0.4953103085937405 => 0.4953985938297469
STATE[0]: 0.49597565076335104 => 0.49591794506999065
STATE[4]: 0.5 => 0.55
STATE[3]: 0.5 => 0.505
STATE[2]: 0.5038068390461955 => 0.5039261551415759
STATE[1]: 0.5046596166281039 => 0.504586270479451
STATE[0]: 0.5041495693099727 => 0.5041932394269205
STATE[4]: 0.5 => 0.45
STATE[3]: 0.5 => 0.495
STATE[2]: 0.4961931609538045 => 0.49607384485842404
STATE[1]: 0.4953985938297469 => 0.4954661189326146
STATE[0]: 0.49591794506999065 => 0.49587276245625306
STATE[6]: 0.5 => 0.55
STATE[5]: 0.5 => 0.505
STATE[4]: 0.5 => 0.5005
STATE[3]: 0.505 => 0.50455
STATE[2]: 0.5039261551415759 => 0.5039885396274183
STATE[1]: 0.504586270479451 => 0.5045264973942478
STATE[0]: 0.5041932394269205 => 0.5042265652236533
STATE[6]: 0.5 => 0.45
STATE[5]: 0.5 => 0.495
STATE[4]: 0.5 => 0.4995
STATE[3]: 0.495 => 0.49545
STATE[2]: 0.49607384485842404 => 0.49601146037258165
STATE[1]: 0.4954661189326146 => 0.4955206530766113
STATE[0]: 0.49587276245625306 => 0.4958375515182889
STATE[4]: 0.5 => 0.55
STATE[3]: 0.50455 => 0.5090950000000001
STATE[2]: 0.5039885396274183 => 0.5044991856646764
STATE[1]: 0.5045264973942478 => 0.5045237662212907
STATE[0]: 0.5042265652236533 => 0.504256285323417
STATE[4]: 0.5 => 0.45
```

Saved successfully! ✕

```
                              => 0.4955008143353235
STATE[1]: 0.4955206530766113 => 0.49551866920248255
STATE[0]: 0.4958375515182889 => 0.4958056632867083
Agent is taking an explore action
STATE[6]: 0.5 => 0.55
```

```
STATE[6]: 0.5 => 0.55
STATE[5]: 0.505 => 0.5095000000000001
STATE[4]: 0.5005 => 0.5014
STATE[3]: 0.5155255000000001 => 0.5141129500000001
STATE[2]: 0.505217954755 => 0.5061074542795
STATE[1]: 0.5045237662212907 => 0.5046821350271116
STATE[0]: 0.504256285323417 => 0.5042988702937865
STATE[6]: 0.5 => 0.45
STATE[5]: 0.495 => 0.4905
STATE[4]: 0.4995 => 0.4986
STATE[3]: 0.4844745 => 0.48588705
STATE[2]: 0.494782045245 => 0.4938925457205
Explore, not updating state value function
STATE[0]: 0.4958056632867083 => 0.4957769638782857
10 turns, player 1 win 1.00, player 2 win 0.00
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:s
-------------
| 0 | 0 | 0 |
-------------
| x | * | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:q
-------------
| * | 0 | 0 |
-------------
| x | * | 0 |
-------------
| 0 | 0 | x |
-------------
Input your position:w
-------------
| * | * | 0 |
-------------
| x | * | x |
-------------
| 0 | 0 | x |
-------------
Input your position:e
You win!
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
```

Saved successfully!   ✕

```
-------------------------------------------------------------------------
KeyboardInterrupt                          Traceback (most recent call last)
/usr/local/lib/python3.6/dist-packages/ipykernel/kernelbase.py in
```

```
        _input_request(self, prompt, ident, parent, password)
    728                 try:
--> 729                     ident, reply = self.session.recv(self.stdin_socket, 0)
    730                 except Exception:
```

⌃⌄ 8 frames

```
zmq/backend/cython/socket.pyx in zmq.backend.cython.socket.Socket.recv()


zmq/backend/cython/socket.pyx in zmq.backend.cython.socket.Socket.recv()


zmq/backend/cython/socket.pyx in zmq.backend.cython.socket._recv_copy()


KeyboardInterrupt:


During handling of the above exception, another exception occurred:


KeyboardInterrupt                         Traceback (most recent call last)
/usr/local/lib/python3.6/dist-packages/ipykernel/kernelbase.py in
_input_request(self, prompt, ident, parent, password)
    732                 except KeyboardInterrupt:
    733                     # re-raise KeyboardInterrupt, to truncate traceback
--> 734                     raise KeyboardInterrupt
```

```python
def play_and_train():
    while True:
        player1 = HumanPlayer()
        player2 = Player(epsilon=0.01)
        judger = Judger(player1, player2)
        winner = judger.play()
        if winner == player2.symbol:
            print("You lose!")
        elif winner == player1.symbol:
            print("You win!")
        else:
            print("It is a tie!")
        player2.backup()
        judger.reset()


if __name__ == '__main__':
    play()
```

Saved successfully! ✕

```
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:s
-------------
| 0 | 0 | 0 |
-------------
| 0 | * | 0 |
-------------
| 0 | x | 0 |
-------------
Input your position:a
-------------
| x | 0 | 0 |
-------------
| * | * | 0 |
-------------
| 0 | x | 0 |
-------------
Input your position:d
You win!
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:s
-------------
| 0 | 0 | 0 |
-------------
| 0 | * | 0 |
-------------
| 0 | x | 0 |
-------------
Input your position:a
-------------
| 0 | 0 | 0 |
-------------
| * | * | x |
-------------
| 0 | x | 0 |
-------------
Input your position:q
-------------
| * | 0 | x |
```

Saved successfully!   ×

```
| 0 | x | 0 |
-------------
Input your position:z
```

```
You win!
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:s
-------------
| 0 | 0 | 0 |
-------------
| 0 | * | 0 |
-------------
| 0 | 0 | x |
-------------
Input your position:a
-------------
| 0 | 0 | 0 |
-------------
| * | * | 0 |
-------------
| x | 0 | x |
-------------
Input your position:d
You win!
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:s
-------------
| 0 | x | 0 |
-------------
| 0 | * | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:a
-------------
| 0 | x | x |
-------------
| * | * | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:d
You win!
-------------
```

Saved successfully! ✕

```
| 0 | 0 | 0 |
-------------
```

```
        Input your position:s
        -------------
        | 0 | 0 | 0 |
        -------------
        | 0 | * | 0 |
        -------------
        | 0 | 0 | x |
        -------------
        Input your position:a
        -------------
        | 0 | 0 | 0 |
        -------------
        | * | * | x |
        -------------
        | 0 | 0 | x |
        -------------
        Input your position:q
        You lose!
        -------------
        | 0 | 0 | 0 |
        -------------
        | 0 | 0 | 0 |
        -------------
        | 0 | 0 | 0 |
        -------------
        Input your position:s
        -------------
        | 0 | 0 | 0 |
        -------------
        | 0 | * | 0 |
        -------------
        | 0 | x | 0 |
        -------------
        Input your position:a
        -------------
        | 0 | 0 | x |
        -------------
        | * | * | 0 |
        -------------
        | 0 | x | 0 |
        -------------
        Input your position:d
        You win!
        -------------
        | 0 | 0 | 0 |
        -------------
        | 0 | 0 | 0 |
        -------------
        | 0 | 0 | 0 |
        -------------
        Input your position:s
```

Saved successfully!   ×

```
        | x | * | 0 |
        -------------
        | 0 | 0 | 0 |
        -------------
```

```
-------------
Input your position:w
-------------
| 0 | * | 0 |
-------------
| x | * | x |
-------------
| 0 | 0 | 0 |
-------------
Input your position:x
You win!
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:s
-------------
| 0 | 0 | 0 |
-------------
| x | * | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:w
-------------
| 0 | * | 0 |
-------------
| x | * | 0 |
-------------
| x | 0 | 0 |
-------------
Input your position:x
You win!
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:s
-------------
| x | 0 | 0 |
-------------
| 0 | * | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:a
```

Saved successfully! ×

```
| * | * | 0 |
-------------
| x | 0 | 0 |
```

```
-------------
Input your position:d
You win!
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:s
-------------
| 0 | 0 | 0 |
-------------
| x | * | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:w
-------------
| x | * | 0 |
-------------
| x | * | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:x
You win!
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:s
-------------
| 0 | 0 | 0 |
-------------
| 0 | * | 0 |
-------------
| x | 0 | 0 |
-------------
Input your position:a
-------------
| 0 | x | 0 |
-------------
| * | * | 0 |
-------------
| x | 0 | 0 |
-------------
Input your position:d
```

Saved successfully! ✕

```
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
```

```
| 0 | 0 | 0 |
-------------
Input your position:s
-------------
| 0 | 0 | 0 |
-------------
| 0 | * | 0 |
-------------
| x | 0 | 0 |
-------------
Input your position:a
-------------
| x | 0 | 0 |
-------------
| * | * | 0 |
-------------
| x | 0 | 0 |
-------------
Input your position:d
You win!
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:s
-------------
| 0 | 0 | 0 |
-------------
| 0 | * | 0 |
-------------
| 0 | 0 | x |
-------------
Input your position:a
-------------
| x | 0 | 0 |
-------------
| * | * | 0 |
-------------
| 0 | 0 | x |
-------------
Input your position:d
You win!
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
```

Saved successfully!   ✕

```
| 0 | 0 | 0 |
-------------
| x | * | 0 |
```

```
-------------
| 0 | 0 | 0 |
-------------
Input your position:w
-------------
| 0 | * | 0 |
-------------
| x | * | x |
-------------
| 0 | 0 | 0 |
-------------
Input your position:x
You win!
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:s
-------------
| 0 | 0 | 0 |
-------------
| 0 | * | 0 |
-------------
| 0 | x | 0 |
-------------
Input your position:a
-------------
| 0 | x | 0 |
-------------
| * | * | 0 |
-------------
| 0 | x | 0 |
-------------
Input your position:d
You win!
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:s
-------------
| 0 | 0 | 0 |
-------------
| 0 | * | 0 |
-------------
| 0 | x | 0 |
```

```
-------------
| 0 | 0 | x |
-------------
| * | * | 0 |
```

```
-------------
| 0 | x | 0 |
-------------
Input your position:d
You win!
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:s
-------------
| 0 | 0 | 0 |
-------------
| 0 | * | 0 |
-------------
| x | 0 | 0 |
-------------
Input your position:a
-------------
| 0 | 0 | 0 |
-------------
| * | * | 0 |
-------------
| x | 0 | x |
-------------
Input your position:d
You win!
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:s
-------------
| 0 | 0 | 0 |
-------------
| 0 | * | 0 |
-------------
| x | 0 | 0 |
-------------
Input your position:a
-------------
| 0 | 0 | 0 |
-------------
| * | * | 0 |
-------------
| x | x | 0 |
-------------
You win!
-------------
| 0 | 0 | 0 |
-------------
```

Saved successfully!   ✕

```
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:s
-------------
| x | 0 | 0 |
-------------
| 0 | * | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:a
-------------
| x | x | 0 |
-------------
| * | * | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:d
You win!
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:s
-------------
| 0 | 0 | x |
-------------
| 0 | * | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:a
-------------
| 0 | x | x |
-------------
| * | * | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:d
You win!
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
```

Saved successfully! ✕

```
Input your position:s
-------------
| 0 | x | 0 |
```

```
-------------
| 0 | * | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:a
-------------
| 0 | x | 0 |
-------------
| * | * | x |
-------------
| 0 | 0 | 0 |
-------------
Input your position:e
-------------
| 0 | x | * |
-------------
| * | * | x |
-------------
| 0 | 0 | x |
-------------
Input your position:z
You win!
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:s
-------------
| 0 | 0 | x |
-------------
| 0 | * | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:a
-------------
| 0 | x | x |
-------------
| * | * | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:d
You win!
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
```

Saved successfully!   ✕

```
Input your position:s
-------------
| 0 | 0 | x |
```

```
 -------------
| 0 | * | 0 |
 -------------
| 0 | 0 | 0 |
 -------------
Input your position:a
 -------------
| x | 0 | x |
 -------------
| * | * | 0 |
 -------------
| 0 | 0 | 0 |
 -------------
Input your position:d
You win!
 -------------
| 0 | 0 | 0 |
 -------------
| 0 | 0 | 0 |
 -------------
| 0 | 0 | 0 |
 -------------
Input your position:s
 -------------
| x | 0 | 0 |
 -------------
| 0 | * | 0 |
 -------------
| 0 | 0 | 0 |
 -------------
Input your position:a
 -------------
| x | 0 | 0 |
 -------------
| * | * | 0 |
 -------------
| 0 | 0 | x |
 -------------
Input your position:d
You win!
 -------------
| 0 | 0 | 0 |
 -------------
| 0 | 0 | 0 |
 -------------
| 0 | 0 | 0 |
 -------------
Input your position:s
 -------------
| 0 | 0 | 0 |
 -------------
| 0 | * | 0 |
 -------------
Input your position:a
 -------------
| 0 | 0 | 0 |
```

Saved successfully!   ✕

```
| 0 | 0 | 0 |
-------------
| * | * | x |
-------------
| x | 0 | 0 |
-------------
Input your position:w
-------------
| 0 | * | 0 |
-------------
| * | * | x |
-------------
| x | x | 0 |
-------------
Input your position:e
You lose!
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:s
-------------
| 0 | 0 | 0 |
-------------
| 0 | * | 0 |
-------------
| 0 | 0 | x |
-------------
Input your position:a
-------------
| 0 | 0 | 0 |
-------------
| * | * | 0 |
-------------
| x | 0 | x |
-------------
Input your position:d
You win!
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:s
-------------
| 0 | 0 | 0 |
-------------
```

Saved successfully!   ✕

```
-------------
Input your position:a
-------------
```

```
| 0 | 0 | x |
-------------
| * | * | 0 |
-------------
| 0 | 0 | x |
-------------
Input your position:d
You win!
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:s
-------------
| 0 | 0 | x |
-------------
| 0 | * | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:a
-------------
| 0 | 0 | x |
-------------
| * | * | 0 |
-------------
| x | 0 | 0 |
-------------
Input your position:d
You win!
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:s
-------------
| x | 0 | 0 |
-------------
| 0 | * | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:a
-------------
| x | 0 | x |
-------------
```

```
| 0 | 0 | x |
-------------
Input your position:d
You win!
```

```
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:s
-------------
| 0 | 0 | 0 |
-------------
| 0 | * | 0 |
-------------
| 0 | x | 0 |
-------------
Input your position:a
-------------
| 0 | 0 | 0 |
-------------
| * | * | 0 |
-------------
| x | x | 0 |
-------------
Input your position:d
You win!
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:s
-------------
| 0 | 0 | 0 |
-------------
| 0 | * | 0 |
-------------
| x | 0 | 0 |
-------------
Input your position:a
-------------
| 0 | 0 | x |
-------------
| * | * | 0 |
-------------
| x | 0 | 0 |
-------------
Input your position:d
You win!
-------------
| 0 | 0 | 0 |
-------------
```

Saved successfully!   ✕

```
-------------
| 0 | 0 | 0 |
-------------
Input your position:s
```

```
Input your position:s
-------------
| 0 | 0 | 0 |
-------------
| 0 | * | x |
-------------
| 0 | 0 | 0 |
-------------
Input your position:w
-------------
| 0 | * | 0 |
-------------
| 0 | * | x |
-------------
| x | 0 | 0 |
-------------
Input your position:x
You win!
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:s
-------------
| x | 0 | 0 |
-------------
| 0 | * | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:a
-------------
| x | 0 | 0 |
-------------
| * | * | 0 |
-------------
| x | 0 | 0 |
-------------
Input your position:d
You win!
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:s
-------------
```

```
-------------
| 0 | 0 | 0 |
-------------
```

```
Input your position:w
-------------
| 0 | * | 0 |
-------------
| x | * | 0 |
-------------
| 0 | 0 | x |
-------------
Input your position:x
You win!
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:s
-------------
| 0 | 0 | x |
-------------
| 0 | * | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:a
-------------
| 0 | 0 | x |
-------------
| * | * | x |
-------------
| 0 | 0 | 0 |
-------------
Input your position:c
-------------
| 0 | 0 | x |
-------------
| * | * | x |
-------------
| x | 0 | * |
-------------
Input your position:w
-------------
| 0 | * | x |
-------------
| * | * | x |
-------------
| x | x | * |
-------------
Input your position:q
You win!
-------------
```

Saved successfully!

```
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
```

```
Input your position:s
-------------
| 0 | 0 | 0 |
-------------
| 0 | * | 0 |
-------------
| x | 0 | 0 |
-------------
Input your position:a
-------------
| x | 0 | 0 |
-------------
| * | * | 0 |
-------------
| x | 0 | 0 |
-------------
Input your position:d
You win!
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:s
-------------
| x | 0 | 0 |
-------------
| 0 | * | 0 |
-------------
| 0 | 0 | 0 |
-------------
Input your position:a
-------------
| x | 0 | 0 |
-------------
| * | * | x |
-------------
| 0 | 0 | 0 |
-------------
Input your position:d
```

```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
<ipython-input-16-6e79b3c33344> in <module>()
      17
      18 if __name__ == '__main__':
---> 19     play()
```

At the beginning the agent is justing making random moves and thus it will always lose. But starting
from some point it starts blocking my moves, not by accident but intentionally since it could block

```
                            te, is_end = all_states[next_state_hash]
      167                   self.p1.set_state(current_state)
      168                   self.p2.set_state(current_state)
```

`KeyError: 3631.0`

SEARCH STACK OVERFLOW

Saved successfully! ✕