CSE 4309 Assignment 7

Inshaad Merchant - 1001861293

Due date: 21 November 2024

Task 1:

value_iteration('environment2.txt', -0.04, 1, 20)

```
utilities:
 0.812  0.868  0.918  1.000
 0.762  0.000  0.660 -1.000
 0.705  0.655  0.611  0.388

policy:
> > > o
^ x ^ o
^ < < <
```

value_iteration('environment2.txt', -0.04, 0.9, 20)

```
utilities:
 0.509  0.650  0.795  1.000
 0.399  0.000  0.486 -1.000
 0.296  0.254  0.345  0.130

policy:
> > > o
^ x ^ o
^ > ^ <
```

Task 2:

The reward of non-terminal states in chess for reinforcement learning algorithms would be kept something very low (close to zero but slightly negative) because this helps the model build a more comprehensive understanding of different game states and strategies. Also, Lower costs for state transitions enable the model to explore longer move sequences without being overly penalized. Thus, users can execute many complex strategies that require execution of more moves.

The discount factor will be kept high (close to 1) during the play to allow the user to take time to explore various states thoroughly and build a comprehensive understanding of different game positions.

Task 3:

Part a:

To Calculate the utility of state (2,2),
We assume that each action:

- Succeeds with probability 0.8
- Has a probability 0.2 of moving in a direction that differs by 90 degrees from the intended direction

We start at the state (2,2) so the only options of moving from there are up, down, right/left.
Case 1: Moving Up
If we move up, this action will succeed with a probability of 0.8.
If we move right or left, it touches the wall, rebounds and stays in position so the probability of that will be 0.1
So U(2,2) = (0.8*$((0.9^0) * -0.04) + ((0.9^1) * 1))) + 0.1(U(2,2)) + 0.1(U(2,2))$
$U(2,2) = 0.8 * (0.86) + 0.2(U(2,2))$
$U(2,2) - 0.2(U(2,2)) = 0.8 * (0.86)$
$0.8 * U(2,2) = 0.8 * (0.86)$
$U(2,2) = 0.86$
Case 2: Moving Down
$U(2,2) = (0.8 * ((0.9^0) * -0.04) + (0.9^1) * -1)) + 0.1(U(2,2)) + 0.1(U(2,2))$
$0.8 * U(2,2) = 0.8 * (-0.94)$
$U(2,2) = -0.94$
Case 3: Moving Right/Left:
$U(2,2) = 0.8 * (-0.04 + 0.9 * U(2,2)) + 0.1 * (-0.04 + 0.9) + 0.1 * (-0.04 - 0.9)$
$U(2,2) = -0.032 + 0.72 * U(2,2) + 0.086 - 0.094$
$0.28 * U(2,2) = -0.04$
$U(2,2) = -0.143$

Part b:

To find the range of values for r where "up" action is not optimal, we need to rewrite our equations from part a replacing -0.04 with r:

Case 1: Moving Up

$U(2,2) = 0.8*(r + 0.9*1) + 0.1*U(2,2) + 0.1*U(2,2)$
$U(2,2) = 0.8*(r + 0.9) + 0.2*U(2,2)$
$0.8*U(2,2) = 0.8*(r + 0.9)$
$U(2,2) = r + 0.9$

Case 2: Moving Down

$U(2,2) = 0.8*(r + 0.9*(-1)) + 0.1*U(2,2) + 0.1*U(2,2)$
$U(2,2) = 0.8*(r - 0.9) + 0.2*U(2,2)$
$U(2,2) = r - 0.9$

Case 3: Moving Right/Left:

$U(2,2) = 0.8*(r + 0.9*U(2,2)) + 0.1*(r + 0.9) + 0.1*(r - 0.9)$
$U(2,2) = 0.8r + 0.72*U(2,2) + 0.1r + 0.09 + 0.1r - 0.09$
$0.28*U(2,2) = r$
$U(2,2) = r/0.28$

For up to not be optimal, at least one other action must give greater or equal utility:

Consider: Down $\geq$ Up : $r - 0.9 \geq r + 0.9$

This is never true

Consider: Right/Left $\geq$ Up

$r/0.28 \geq r + 0.9$
$r/0.28 - r \geq 0.9$
$r(1/0.28 - 1) \geq 0.9$
$r(3.57 - 1) \geq 0.9$
$2.57r \geq 0.9$
$r \geq 0.35$

Therefore, the range of values for r where the "up" action is not optimal is $r \geq 0.35$. Then left/right will be the best action to take.