

Name: Sujan Shrestha
UTA ID: 1001752468

Assignment 7 Written Task

Task 1

Output for **python3 value_iteration.py environment2.txt -0.04 1 20** is
utilities:

```
0.812 0.868 0.918 1.000
0.762 0.000 0.660 -1.000
0.705 0.655 0.611 0.387
```

policy:

```
> > > o
^ X ^ o
^ < < <
```

Output for **python3 value_iteration.py environment2.txt -0.04 0.9 20** is
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

In playing chess, I would assign the value for reward of non-terminal states as 0 or as lower value as possible. This is because there are so many possible states to be accessed, thus the higher penalty can be ignored in taking such large number of the states before the game reaches to terminal state.

For the discount factor gamma, I would choose 1. This is because it leads in reaching terminal states and winning the game rather than focusing on the immediate rewards or penalty in the intermediate steps.

Task3

Part a: As (2,1) and (2,3) are blocked and we are in the state (2,2). From there we can go up to (3,2) with terminal reward +1 and down to (1,2) with terminal reward -1.

Utility in upper case is

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

$$U(2,2)-0.18(U(2,2))=r+0.72$$

$$U(2,2)=(r+0.72)/0.82$$

$$U(2,2) = (-0.04+0.72)/0.82 = 0.82927$$

Utility on left and right side is given by

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

$$U(2,2)-0.72(U(2,2))=r$$

$$U(2,2)=r/0.28$$

$$U(2,2)= 3.571r$$

$$U(2,2)=3.571*(-0.04) = -0.1428$$

Utility on down will obviously not be optimal as the reward is -1.

Therefore, utility of the state $U(2,2)$ is the max value among 4 cases that is 0.82927 in UPPER CASE making it the optimal case.

Part b:

Utility in up side is given from part a

$$U(2,2)=(r+0.72)/0.82$$

Utility on left and right side is given from part a

$$U(2,2)= 3.571r$$

Utility on down will obviously not be optimal as the reward is -1.

So, next optimal could be left side or right side for which the utility value of left or right should be more than that of utility of upper case,

$$3.571r > (r+0.72)/0.82$$

$$2.92822r > r+0.72$$

$$1.92822r > 0.72$$

$$r > 0.3734$$

Therefore, the reward value greater than 0.3734 will change the optimal solution from being up to left or right.