Team 63

Abhinav Gupta - 20171059
Praffullitt Jain - 20171142

PART B

X = 63

1)
The discount factor describes the preference of an agent for current rewards over future rewards. When γ is close to 0, rewards in the distant future are viewed as insignificant. When γ is 1, discounted rewards are exactly equivalent to additive rewards, so additive rewards are a special case of discounted rewards.

1.a
gamma = 0.1, step cost = -6.3

Utility Map:

6.3		-1.343	63.0
-5.924	-6.912	-6.494	-1.343
-6.912	-6.993		-6.549
-6.993	-6.993	-12.6	-6.993

Policy Table:

		East	
North	West	East	North
North	West		North
North	West		East

# 1.b
gamma = 0.99, step cost = -6.3

Utility Map:

6.3		53.422	63.0
24.878	36.257	45.837	53.422
18.568	26.919		44.901
10.962	14.858	-12.6	31.092

Policy Table:

		East	
East	East	East	North
East	North		North
North	North		North

Comparing 1a and 1b:

The values in 1a utility map are negative because value of gamma is very less (close to 0), so it is not taking into consideration, the distant future values. Whereas in 1b, the gamma value is close to 1, so the future values are taking into considering, resulting in positive values in the utility map.

# 2)
gamma = 0.99

2.a Step cost: X = 63
Since the step cost is positive, instead of the agent approaching the end states; it will just keep looping around because it is still able to maximise its reward/utility.

Utility Map:

6.3		6299.989	63.0
6299.9899	6299.9899	6299.9899	6299.9899
6299.9899	6299.9899		6299.9899
6299.9899	6299.9899	6299.9899	6299.9899

Policy Table:

		West	
South	West	West	South
West	West		West
West	West		East

2.b Step cost X = -12.6 (-X/5)

Utility Map:

6.3		44.716	63.0
-4.295	11.993	30.235	44.716
-18.359	-5.459		28.447
-32.473	-21.386	-12.6	9.637

Policy Table:

		East	
East	East	East	North
North	North		North
North	North		North

2.c step cost X = -15.75

Utility Map:

6.3		40.362	63.0
-11.948	-0.051	22.434	40.362
-30.274	-20.852		20.221
-47.482	-30.847	-12.6	-1.090

Policy Table:

		East	
North	East	East	North
North	North		North
North	East		North

Comparing 2b and 2c:
The absolute value of step cost in 2c is greater than 2b. In case 2b, when the agent reached (1,0), it can either go east to the maximum reward cell or just go north to an end state with lesser reward. It chooses to go east however, because the step cost is lesser.
Whereas in 2c, the agent chooses to reach the end state as soon as possible because traversing for more time in the grid would cost it more.

2.d step cost X = -63

Extreme case of 2a. Since the step cost is negative, the agent will try to find the easiest way to reach the end state because moving around the grid is going to cost it a lot.

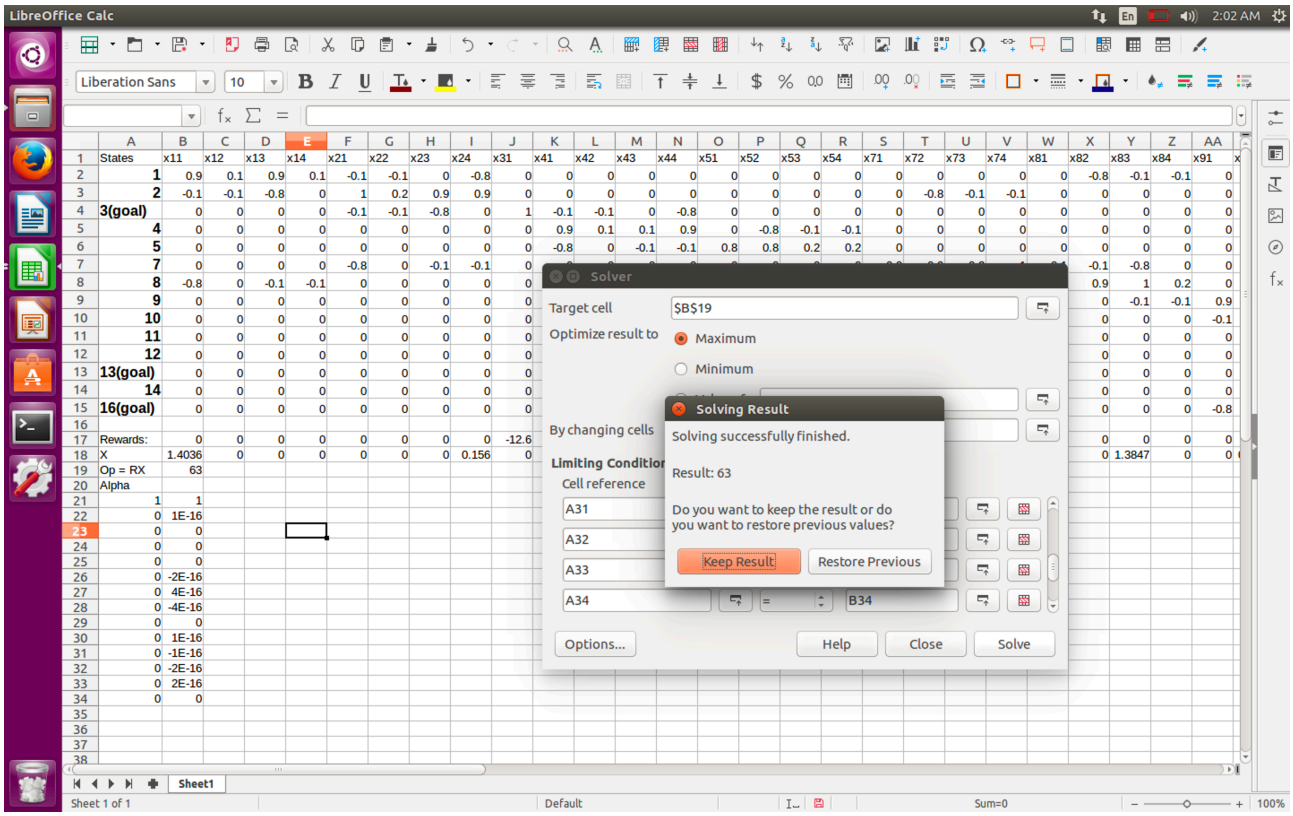
Utility Map:

6.3		-24.936	63.0
-82.122	-161.430	-94.581	-24.936
-161.430	-175.842		-103.179
-175.842	-100.319	-12.6	-92.335

Policy Table:

		East	
North	West	East	North
North	South		North
East	East		West

PART C:


LibreOffice Calc
Solver
Target cell
B19
Optimize result to
Maximum
Minimum
By changing cells
Limiting Condition
Cell reference
A31
A32
A33
A34
B34
Options...
Help
Close
Solve
Solving Result
Solving successfully finished.
Result: 63
Do you want to keep the result or do you want to restore previous values?
Keep Result
Restore Previous
Sheet1
Sheet 1 of 1
Default
Sum=0
100%