Team 64

Roll Number - 201501051 (Tirth Maniar)
- 201501066 (Moin Moti)

Value Iteration

Matrix till convergence:

(Delta = 3.2)

Iteration 0: Initial Board

W	W	64	W
0.000000	0.000000	0.000000	0.000000
0.000000	-64	W	0.000000
0.000000	0.000000	0.000000	0.000000

Iteration 1:

W	W	64	W
-3.200000	-3.200000	48.000000	-3.200000
-3.200000	-64	W	-3.200000
-3.200000	-3.200000	-3.200000	-3.200000

Max Change = 48

Iteration 2:

W	W	64	W
-6.400000	28.480000	47.360000	34.560000

-6.400000	-64	W	-6.400000
-6.400000	-6.400000	-6.400000	-6.400000

Max Change = 37.76

Iteration 3:

W	W	64	W
18.304000	31.136000	54.304000	37.504000
-9.600000	-64	W	23.168000
-9.600000	-9.600000	-9.600000	-9.600000

Max Change = 29.568

Iteration 4:

W	W	64	W
22.579200	36.956800	54.864000	46.310400
4.083200	-64	W	31.436800
-12.800000	-12.800000	-12.800000	13.414400

Max Change = 23.0144

Iteration 5:

W	W	64	W
29.031680	37.986880	56.326720	48.465920
8.871680	-64	W	40.135680
-2.493440	-16.000000	4.971520	22.010880

Max Change = 17.7715

Iteration 6:

W	W	64	W
30.979840	39.260064	56.645280	50.721536
14.512512	-64	W	43.599872
2.048000	-7.222784	15.403008	31.606784

Max Change = 10.4315

Iteration 7:

W	W	64	W
32.757286	39.642230	56.998160	51.548365
16.635123	-64	W	46.097203
7.892531	2.000128	25.166029	36.380877

Max Change = 9.76302

Iteration 8:

W	W	64	W
33.453025	39.962751	57.119060	52.163085
18.269341	-64	W	47.258132
11.097364	10.732836	30.937907	39.832453

Max Change = 8.73271

Iteration 9:

W	W	64	W
33.942438	40.091523	57.212584	52.437369
18.989354	-64	W	47.982094
13.598493	16.223609	34.853544	41.683542

Max Change = 5.49077

Iteration 10:

W	W	64	W
34.166397	40.179219	57.252889	52.612013
19.452885	-64	W	48.346314
14.973694	19.905196	37.117542	42.839384

Max Change = 3.68159

Iteration 11:

W	W	64	W
34.305304	40.220233	57.279123	52.698144
19.678406	-64	W	48.558873
16.166815	22.084554	38.495016	43.472744

Max Change = 2.17936 < Delta(3.2)

Results for Delta = 0:

W	W	64	W
34.462000	40.273973	57.308219	52.808219
19.966223	-64	W	48.808219
21.029901	25.162861	40.308219	44.308219

Expected Reward:

The final expected reward is 16.166815
(21.029901 if delta = 0)

Optimal Policy for each state:

W	W	64	W
Right	Right	Above	Left
Above	-64	W	Above
Right	Right	Right	Above

Optimal Path from start to end:

W	W	64	W
-	-	Above	Left
-	-64	W	Above
Right	Right	Right	Above

Linear Programming

Corresponding States:

W	W	12	W
8	9	10	11
5	6	W	7
1	2	3	4

Actions:

1 - North
2 - East
3 - South
4 - West

values of x:

State,Action Pair	Value of X
1,1	0
1,2	1.111111111
1,3	0
1,4	0
2,1	0
2,2	0.987654321
2,3	0
2,4	0
3,1	0
3,2	1.111111111
3,3	0
3,4	0
4,1	0.987654321
4,2	0
4,3	0
4,4	0
5,1	0.1369863014
5,2	0

5,3	0
5,4	0
6,5	0.1352974292
7,1	1.127999833
7,2	0
7,3	0
7,4	0
8,1	0
8,2	0.1217656012
8,3	0
8,4	0
9,1	0
9,2	0.2283336693
9,3	0
9,4	0
10,1	1.080878214
10,2	0
10,3	0
10,4	0
11,1	0
11,2	0
11,3	0
11,4	1.122764098
12,5	0.8647025708

Expected Reward:

21.02990161

Description of why the rewards match/don't match:

We try to maximize the utility/reward in both the methods of solving the MDP, so they would both end up achieving the same result if we try make them as accurate as possible. In VI, the reward in the start state is the utility of selecting the best paths possible to the terminal states. In LP, the paths we get match the ones in VI. Reward in LP is the summation of the reward*x for each state, action pair. This will correspond to the value we get in VI if we assume a small delta, since a large delta would not allow enough iterations so that our VI could not spread out enough and thus does not approximates the utilities of different states enough times. So if we use a delta not near 0, the values in VI and LP might not match as in our case, but on using delta=0 the rewards match in both of them.