

Robot Learning Exercise 3

Member 1

Name: Hojun Lim

Mat. No: 3279159

Member 2

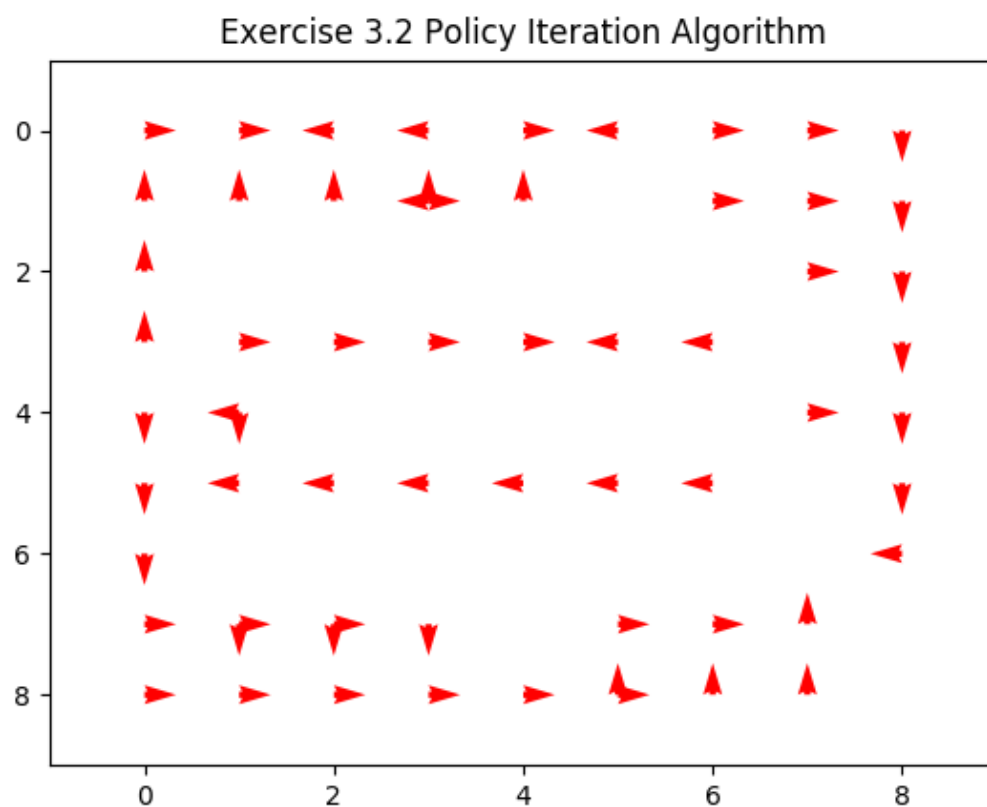
Name: Kajaree Das

Mat. No: 3210311

Solution for 3.1:

0	-40.63	-65.95	-81.72	-90.46	-98.14	-101.22	-82.59	-57.64	-30.43
1	-40.77	-79.81	-96.57	-107.0	-115.4	X	-98.63	-67.07	-34.87
2	-39.09	X	X	X	X	X	X	-81.94	-34.2
3	-38.96	-70.06	-114.24	-131.77	-138.33	-146.42	-156.57	X	-29.28
4	-37.11	-74.07	X	X	X	X	X	-97.02	-26.62
5	-33.39	-89.59	-127.03	-142.48	-149.81	-155.58	-164.93	X	-14.07
6	-24.77	X	X	X	X	X	X	G	-7.18
7	-20.63	-14.48	-12.1	-18.26	X	-19.86	-15.09	-6.54	X
8	-23.43	-18.37	-15.93	-20.57	-28.2	-22.18	-20.57	-25.11	X
9									
0	1	2	3	4	5	6	7	8	9

Solution for 3.2:

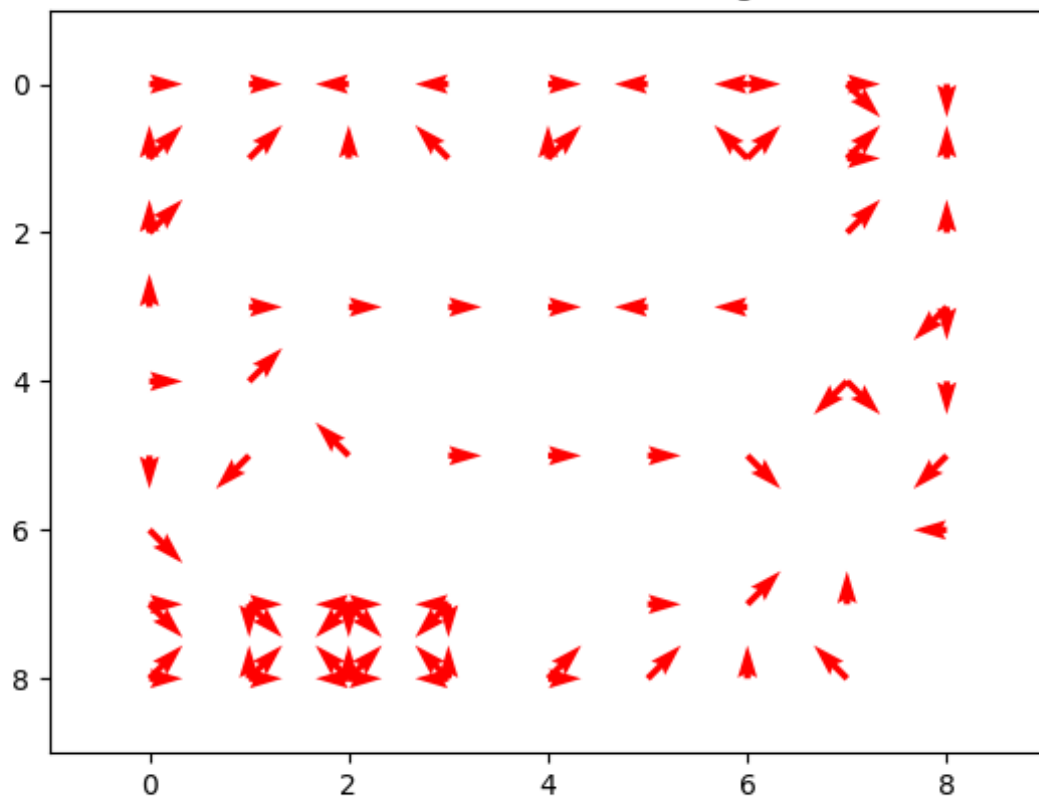


Exercise 3.2 optimal value of all cells

0	49.18	49.18	49.26	49.26	49.18	49.26	53.61	54.01	54.46
1	49.18	49.26	49.26	43.26	49.26	X	48.01	54.46	54.95
2	43.26	X	X	X	X	X	X	54.95	62.17
3	37.86	43.09	49.18	49.18	49.18	49.26	49.26	X	70.19
4	43.11	37.8	X	X	X	X	X	70.19	79.1
5	49.01	43.11	37.8	33.02	28.72	24.84	21.36	X	89.0
6	55.57	X	X	X	X	X	X	G	100.0
7	62.85	64.28	65.87	67.63	X	85.1	89.0	100.0	X
8	64.28	65.87	67.63	69.59	78.43	81.59	85.1	89.0	X
9									
0	1	2	3	4	5	6	7	8	9

Solution for 3.3:

Exercise 3.3 Value Iteration Algorithm



Exercise 3.3 optimal value of all cells									
0	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
1	0.7	0.7	0.7	0.7	0.7	X	0.7	0.7	0.7
2	-0.05	X	X	X	X	X	X	0.7	0.7
3	-0.13	-0.05	0.7	0.7	0.7	0.7	0.7	X	0.02
4	-0.13	-0.05	X	X	X	X	X	1.28	1.28
5	-0.05	-0.05	-0.13	-0.14	0.02	1.28	12.5	X	12.5
6	0.7	X	X	X	X	X	X	G	12.5
7	0.7	0.7	0.7	0.7	X	2.03	12.5	12.5	X
8	0.7	0.7	0.7	0.7	0.85	2.03	2.03	2.03	X
9									

Exercise 3.2 and 3.3 Comparison

Compare to policy iteration algorithm of 3.2, we can clearly see that value iteration algorithm of 3.3 converge much earlier than the algorithm of 3.2 in terms of calculated optimal values. Moreover, as it is allowed to move in diagonal direction in 3.3, the value iteration algorithm choose a diagonal action when it can reach to the higher value states or the goal quicker, instead of traveling around(indirect way) like the optimal policy of 3.2 does.

Exercise 3.3 and 3.4 Comparison

Compare to value iteration algorithm of 3.3, we can clearly see that value iteration algorithm of 3.4 converge much later than the algorithm of 3.3 in terms of calculated optimal values.

Moreover, as it is allowed to move in any desired direction in 3.4 and allowed to drift 45 degrees to the right or left with a probability, the value iteration algorithm chooses suitable action more freely to reach a higher value state or the goal quicker, instead of being bound by only desired or allowed actions like the optimal policy of 3.3 judging by the path our agent takes around the grid.