

# Data Mining Techniques 2013

Tobias Stahl  
10528199

Spyros Michaelides  
10523316

Ioannis Giounous Aivalis  
s2538161

Francesco xxxxxxxx  
xxxxxxxxxx

## 1 Introduction

### 2 Part 1

In the initial section of the assignment an environment is constructed for the simulation of the agent, (the predator), and the prey. The environment is a two dimensional toroidal grid in which if an agent moves off the boundary of a side of the grid, he will appear in the opposite side of the board. The board has a size of 11 by 11 cells.

Both the predator and the prey move with a random policy inclusive of directions NORTH, SOUTH, WEST, EAST along with an option of staying still. While the predator has a 0.2 probability of taking any action, the prey has a probability 0.8 of staying in the same cell, and moves with a probability of only 0.05 in any direction.

The probability for action taken by the prey is changed when the predator will be standing in a neighboring cell, since the prey is not supposed to move into the cell of the predator. This will change the probability of the prey moving to 0.067 toward any of the remaining directions.

In order to find the average of moves it takes the predator to catch the prey a simulation of 100 runs is made. The resulting value for the average is 226 actions with a standard deviation of 289.02

### 3 Part 2

In the second part of the assignment, an iterative policy evaluation is used to determine the value for all possible states, using a discount factor of 0.8. The value of a state is determined by summing the expected reward and

value of all possible next states  $s'$  multiplied by the probability of actually moving into this state  $s'$ .

This is repeated as long as the policy keeps changing. Once this stops, the best policy is found.

The policy evaluation will stop once the change in values is smaller than threshold  $\Theta$ , where  $\Theta$  is 0.00001.

The following images display the values after the policy evaluation of the different states of predator and prey.

### 3.1 State predator(0,0), prey(5,5)

The values of all possible states with the prey beeing in (5,5) are shown in Table 1

	0	1	2	3	4	5	6	7	8	9	10
0	0.004	0.007	0.013	0.022	0.033	0.04	0.033	0.022	0.013	0.007	0.004
1	0.007	0.012	0.025	0.048	0.08	0.102	0.08	0.048	0.025	0.012	0.007
2	0.013	0.025	0.057	0.125	0.237	0.335	0.237	0.125	0.057	0.025	0.013
3	0.022	0.048	0.125	0.316	0.705	1.181	0.705	0.316	0.125	0.048	0.022
4	0.033	0.08	0.237	0.705	1.965	4.455	1.965	0.705	0.237	0.08	0.033
5	0.04	0.102	0.335	1.181	4.455	5.775	4.455	1.181	0.335	0.102	0.04
6	0.033	0.08	0.237	0.705	1.965	4.455	1.965	0.705	0.237	0.08	0.033
7	0.022	0.048	0.125	0.316	0.705	1.181	0.705	0.316	0.125	0.048	0.022
8	0.013	0.025	0.057	0.125	0.237	0.335	0.237	0.125	0.057	0.025	0.013
9	0.007	0.012	0.025	0.048	0.08	0.102	0.08	0.048	0.025	0.012	0.007
10	0.004	0.007	0.013	0.022	0.033	0.04	0.033	0.022	0.013	0.007	0.004

Table 1: Policy evaluation of state predator(0,0), prey(5,5)

### 3.2 State predator(2,3), prey(5,4)

The values of all states with the prey starting in (5,4) are shown in Table 2

### 3.3 State predator(2,10), prey(10,0)

The values of all possible states with the prey beeing in (10,0) are shown in Table 3

	0	1	2	3	4	5	6	7	8	9	10
0	0.007	0.013	0.022	0.033	0.04	0.033	0.022	0.013	0.007	0.004	0.004
1	0.012	0.025	0.048	0.08	0.102	0.08	0.048	0.025	0.012	0.007	0.007
2	0.025	0.057	0.125	0.237	0.335	0.237	0.125	0.057	0.025	0.013	0.013
3	0.048	0.125	0.316	0.705	1.181	0.705	0.316	0.125	0.048	0.022	0.022
4	0.08	0.237	0.705	1.965	4.455	1.965	0.705	0.237	0.08	0.033	0.033
5	0.102	0.335	1.181	4.455	5.775	4.455	1.181	0.335	0.102	0.04	0.04
6	0.08	0.237	0.705	1.965	4.455	1.965	0.705	0.237	0.08	0.033	0.033
7	0.048	0.125	0.316	0.705	1.181	0.705	0.316	0.125	0.048	0.022	0.022
8	0.025	0.057	0.125	0.237	0.335	0.237	0.125	0.057	0.025	0.013	0.013
9	0.012	0.025	0.048	0.08	0.102	0.08	0.048	0.025	0.012	0.007	0.007
10	0.007	0.013	0.022	0.033	0.04	0.033	0.022	0.013	0.007	0.004	0.004

Table 2: Policy evaluation of state predator(2,3), prey(5,4)

	0	1	2	3	4	5	6	7	8	9	10
0	4.455	1.965	0.705	0.237	0.08	0.033	0.033	0.08	0.237	0.705	1.965
1	1.181	0.705	0.316	0.125	0.048	0.022	0.022	0.048	0.125	0.316	0.705
2	0.335	0.237	0.125	0.057	0.025	0.013	0.013	0.025	0.057	0.125	0.237
3	0.102	0.08	0.048	0.025	0.012	0.007	0.007	0.012	0.025	0.048	0.08
4	0.04	0.033	0.022	0.013	0.007	0.004	0.004	0.007	0.013	0.022	0.033
5	0.04	0.033	0.022	0.013	0.007	0.004	0.004	0.007	0.013	0.022	0.033
6	0.102	0.08	0.048	0.025	0.012	0.007	0.007	0.012	0.025	0.048	0.08
7	0.335	0.237	0.125	0.057	0.025	0.013	0.013	0.025	0.057	0.125	0.237
8	1.181	0.705	0.316	0.125	0.048	0.022	0.022	0.048	0.125	0.316	0.705
9	4.455	1.965	0.705	0.237	0.08	0.033	0.033	0.08	0.237	0.705	1.965
10	5.775	4.455	1.181	0.335	0.102	0.04	0.04	0.102	0.335	1.181	4.455

Table 3: Policy evaluation of state predator(2,10), prey(10,0)

### 3.3.1 State predator(10,10), prey(0,0)

The values of all possible states with the prey beeing in (0,0) are shown in Table 4

## 4 Part 3

In section three Policy Iteration has been implemented in order to find an optimal policy. Once the all the states are evaluated using Policy Evaluation (explained in part 2), a policy improvement algorithm is used to find the

	0	1	2	3	4	5	6	7	8	9	10
0	5.775	4.455	1.181	0.335	0.102	0.04	0.04	0.102	0.335	1.181	4.455
1	4.455	1.965	0.705	0.237	0.08	0.033	0.033	0.08	0.237	0.705	1.965
2	1.181	0.705	0.316	0.125	0.048	0.022	0.022	0.048	0.125	0.316	0.705
3	0.335	0.237	0.125	0.057	0.025	0.013	0.013	0.025	0.057	0.125	0.237
4	0.102	0.08	0.048	0.025	0.012	0.007	0.007	0.012	0.025	0.048	0.08
5	0.04	0.033	0.022	0.013	0.007	0.004	0.004	0.007	0.013	0.022	0.033
6	0.04	0.033	0.022	0.013	0.007	0.004	0.004	0.007	0.013	0.022	0.033
7	0.102	0.08	0.048	0.025	0.012	0.007	0.007	0.012	0.025	0.048	0.08
8	0.335	0.237	0.125	0.057	0.025	0.013	0.013	0.025	0.057	0.125	0.237
9	1.181	0.705	0.316	0.125	0.048	0.022	0.022	0.048	0.125	0.316	0.705
10	4.455	1.965	0.705	0.237	0.08	0.033	0.033	0.08	0.237	0.705	1.965

Table 4: Policy evaluation of state predator(10,10), prey(0,0)

best action for the predator to take, regarding the reward to expect in the next state.

In order to do this the best action for each state is chosen by changing the probability of taking this action in this state to 1.0, while the rest of the actions in the state are reduced to a probability of 0.

Because of this change of probabilities the states need to be re-evaluated, leading to an iterative process until the policy is stable.

#### 4.1 Discount Factor ( $\gamma=0.1$ )

For  $\gamma=0.1$  the convergence of policy iteration occurs after 3 iterations. On the other hand the convergence of the policy in value iteration occurs after 11 iterations. Because of the small  $\gamma$  not every state has a value over 0.001 as you can see in Table 5

#### 4.2 Discount Factor ( $\gamma=0.5$ )

For  $\gamma=0.5$  the convergence of policy iteration occurs after 3 iterations. On the other hand the convergence of the policy in value iteration occurs after 31 iterations. The state values of the optimal policy can be seen in Table 6

	0	1	2	3	4	5	6	7	8	9	10
(0)	0	0	0	0	0	0.001	0	0	0	0	0
(1)	0	0	0	0	0.001	0.011	0.001	0	0	0	0
(2)	0	0	0	0.001	0.011	0.111	0.011	0.001	0	0	0
(3)	0	0	0.001	0.011	0.111	1.111	0.111	0.011	0.001	0	0
(4)	0	0.001	0.011	0.111	1.111	11.111	1.111	0.111	0.011	0.001	0
(5)	0.001	0.011	0.111	1.111	11.111	11.111	11.111	1.111	0.111	0.011	0.001
(6)	0	0.001	0.011	0.111	1.111	11.111	1.111	0.111	0.011	0.001	0
(7)	0	0	0.001	0.011	0.111	1.111	0.111	0.011	0.001	0	0
(8)	0	0	0	0.001	0.011	0.111	0.011	0.001	0	0	0
(9)	0	0	0	0	0.001	0.011	0.001	0	0	0	0
(10)	0	0	0	0	0	0.001	0	0	0	0	0

Table 5: Policy Iteration for ( $\gamma=0.1$ ), convergence after 3 iterations.

	0	1	2	3	4	5	6	7	8	9	10
(0)	0.039	0.078	0.156	0.312	0.625	1.25	0.625	0.312	0.156	0.078	0.039
(1)	0.078	0.156	0.312	0.625	1.25	2.5	1.25	0.625	0.312	0.156	0.078
(2)	0.156	0.312	0.625	1.25	2.5	5	2.5	1.25	0.625	0.312	0.156
(3)	0.312	0.625	1.25	2.5	5	10	5	2.5	1.25	0.625	0.312
(4)	0.625	1.25	2.5	5	10	20	10	5	2.5	1.25	0.625
(5)	1.25	2.5	5	10	20	20	20	10	5	2.5	1.25
(6)	0.625	1.25	2.5	5	10	20	10	5	2.5	1.25	0.625
(7)	0.312	0.625	1.25	2.5	5	10	5	2.5	1.25	0.625	0.312
(8)	0.156	0.312	0.625	1.25	2.5	5	2.5	1.25	0.625	0.312	0.156
(9)	0.078	0.156	0.312	0.625	1.25	2.5	1.25	0.625	0.312	0.156	0.078
(10)	0.039	0.078	0.156	0.312	0.625	1.25	0.625	0.312	0.156	0.078	0.039

Table 6: Policy Iteration for ( $\gamma=0.5$ ), convergence after 3 iterations.

### 4.3 Discount Factor ( $\gamma=0.7$ )

For  $\gamma=0.7$  the convergence of policy iteration occurs after 3 iterations. On the other hand the convergence of the policy in value iteration occurs after 60 iterations. The state values of the optimal policy can be seen in Table 7

### 4.4 Discount Factor ( $\gamma=0.9$ )

For  $\gamma=0.9$  the convergence of policy iteration occurs after 3 iterations. On the other hand the convergence of the policy in value iteration occurs after

	0	1	2	3	4	5	6	7	8	9	10
(0)	1.345	1.922	2.745	3.922	5.602	8.003	5.602	3.922	2.745	1.922	1.345
(1)	1.922	2.745	3.922	5.602	8.003	11.433	8.003	5.602	3.922	2.745	1.922
(2)	2.745	3.922	5.602	8.003	11.433	16.333	11.433	8.003	5.602	3.922	2.745
(3)	3.922	5.602	8.003	11.433	16.333	23.333	16.333	11.433	8.003	5.602	3.922
(4)	5.602	8.003	11.433	16.333	23.333	33.333	23.333	16.333	11.433	8.003	5.602
(5)	8.003	11.433	16.333	23.333	33.333	33.333	33.333	23.333	16.333	11.433	8.003
(6)	5.602	8.003	11.433	16.333	23.333	33.333	23.333	16.333	11.433	8.003	5.602
(7)	3.922	5.602	8.003	11.433	16.333	23.333	16.333	11.433	8.003	5.602	3.922
(8)	2.745	3.922	5.602	8.003	11.433	16.333	11.433	8.003	5.602	3.922	2.745
(9)	1.922	2.745	3.922	5.602	8.003	11.433	8.003	5.602	3.922	2.745	1.922
(10)	1.345	1.922	2.745	3.922	5.602	8.003	5.602	3.922	2.745	1.922	1.345

Table 7: Policy Iteration for ( $\gamma=0.7$ ), convergence after 3 iterations.

198 iterations. The state values of the optimal policy can be seen in Table 8

	0	1	2	3	4	5	6	7	8	9	10
(0)	38.742	43.047	47.83	53.144	59.049	65.61	59.049	53.144	47.83	43.047	38.742
(1)	43.047	47.83	53.144	59.049	65.61	72.9	65.61	59.049	53.144	47.83	43.047
(2)	47.83	53.144	59.049	65.61	72.9	81	72.9	65.61	59.049	53.144	47.83
(3)	53.144	59.049	65.61	72.9	81	90	81	72.9	65.61	59.049	53.144
(4)	59.049	65.61	72.9	81	90	100	90	81	72.9	65.61	59.049
(5)	65.61	72.9	81	90	100	100	100	90	81	72.9	65.61
(6)	59.049	65.61	72.9	81	90	100	90	81	72.9	65.61	59.049
(7)	53.144	59.049	65.61	72.9	81	90	81	72.9	65.61	59.049	53.144
(8)	47.83	53.144	59.049	65.61	72.9	81	72.9	65.61	59.049	53.144	47.83
(9)	43.047	47.83	53.144	59.049	65.61	72.9	65.61	59.049	53.144	47.83	43.047
(10)	38.742	43.047	47.83	53.144	59.049	65.61	59.049	53.144	47.83	43.047	38.742

Table 8: Policy Iteration for ( $\gamma=0.9$ ), convergence after 3 iterations.

## 5 Value Iteration

In section 4, value iteration is implemented. The difference between value iteration and policy iteration, is that "Value iteration effectively combines, in each of its sweeps, one sweep of policy evaluation and one sweep of policy improvement." [1]

A plot of all  $\gamma$  between 0 and 1 can be seen in Figure 1.

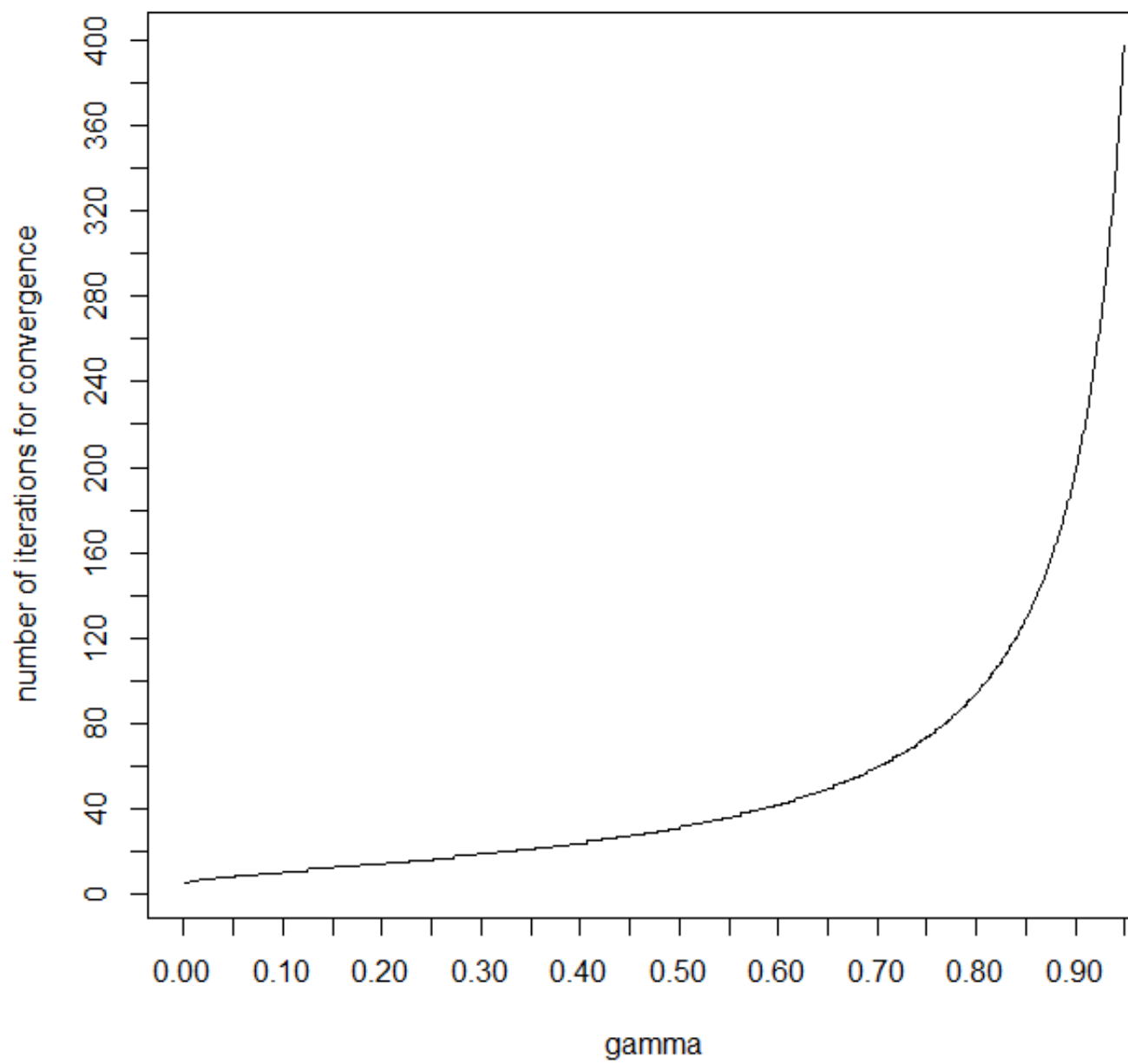


Figure 1: Plot of  $\gamma$  and their number of iterations to converge

## References

- [1] <http://webdocs.cs.ualberta.ca/~sutton/book/ebook/node44.html>