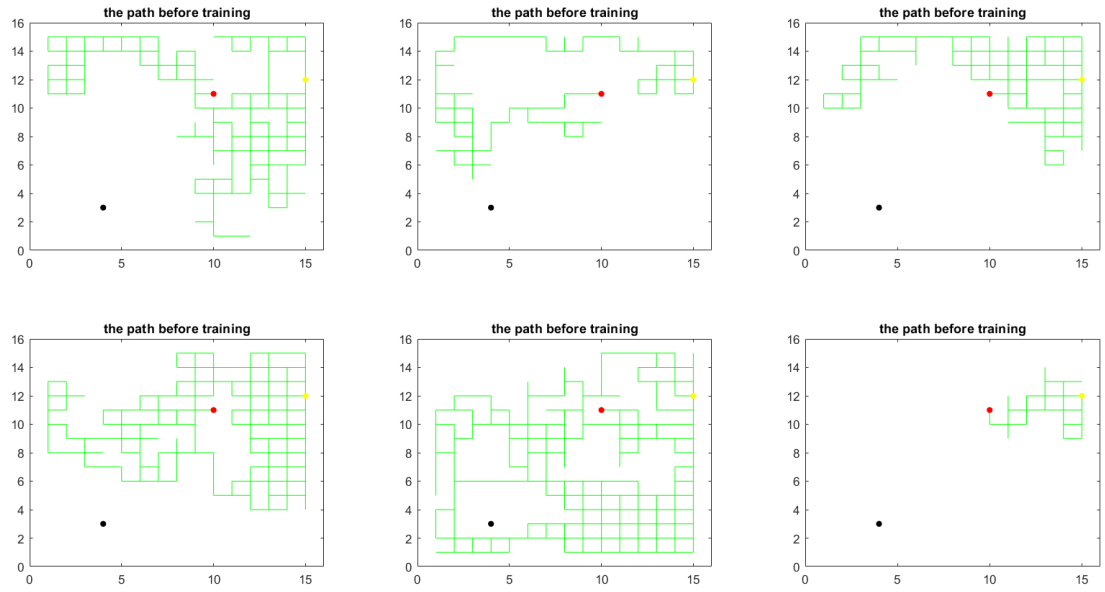
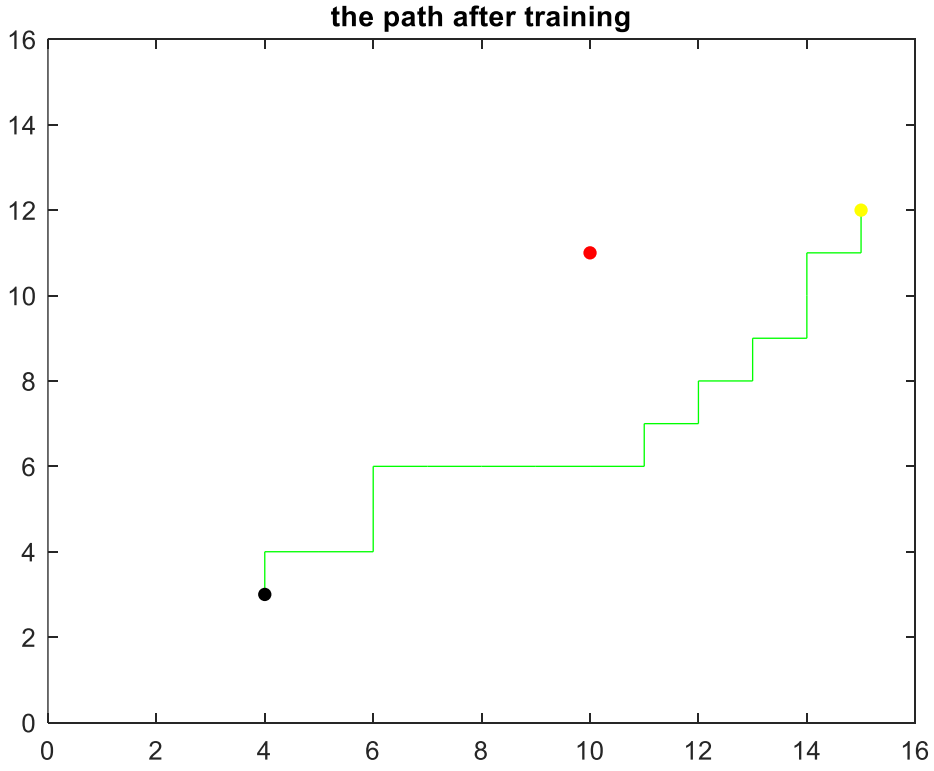
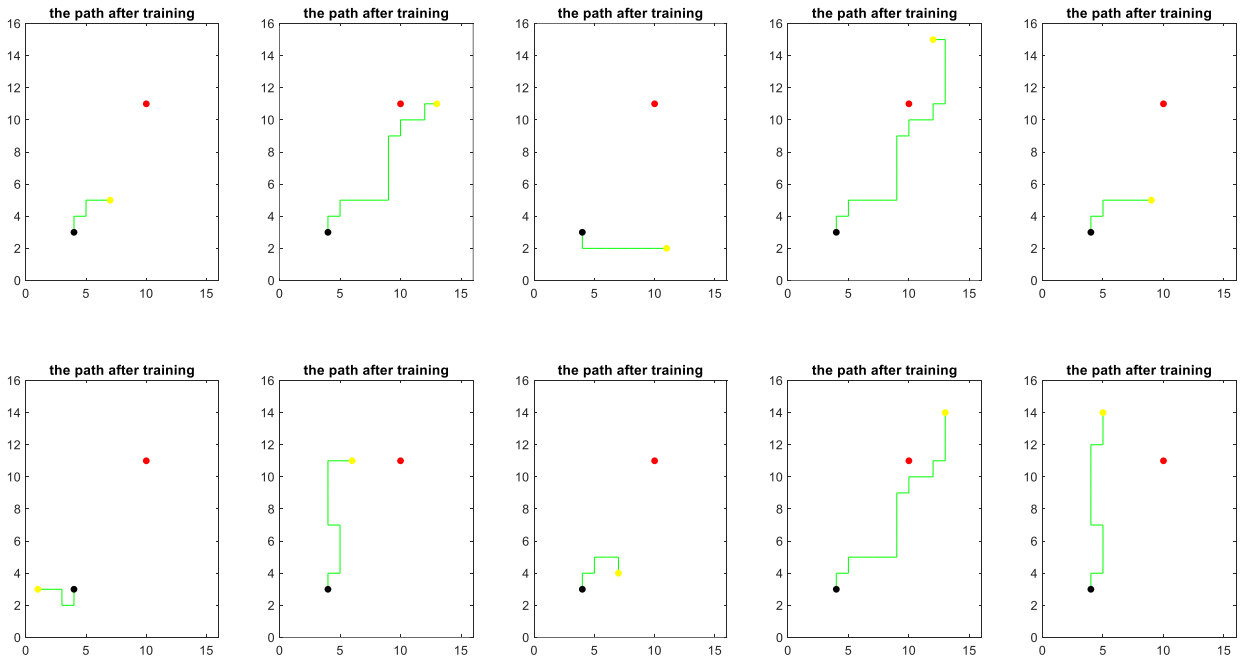
Advanced Topics in Neuroscience *problem set #6 (reinforcement learning)*
Instructor: Dr. Ali Ghazizadeh *Name: Sana Aminnaji*
Student ID: 98104722

1. In this section we first show the path that mouse traverses in the first trials before learning:

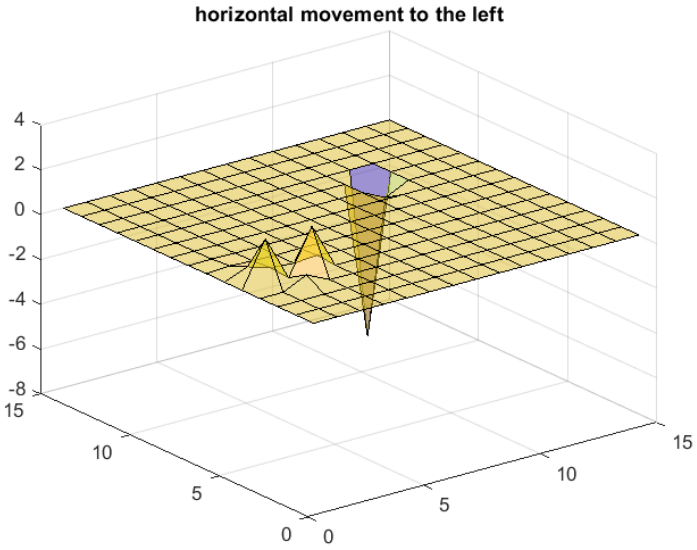


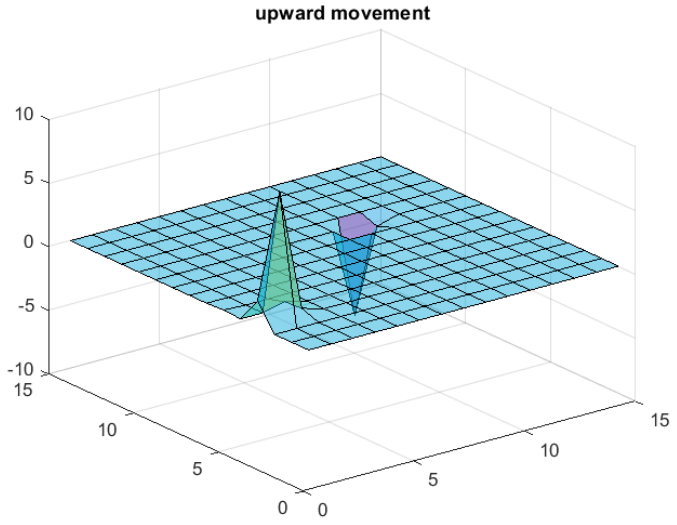
Next, we try to show how the mouse behave after learning, starting from the same initial point:

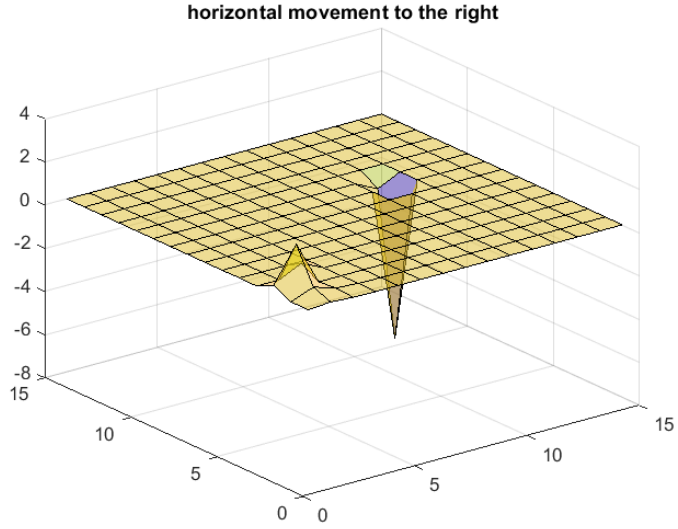


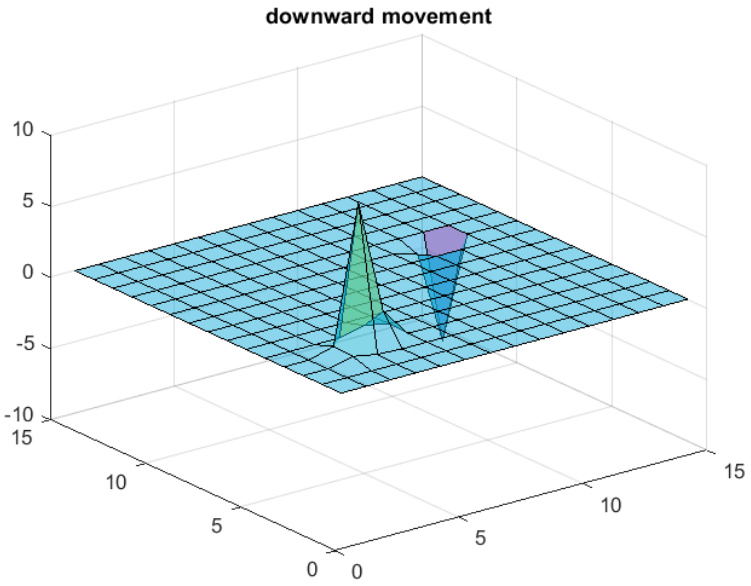


2. We tried to learn the agent using Q-learning approach so we can have the Q(s, a) amount for each set of state and actions. We can show these values in four different plots, indicating the data on the four existing directions:



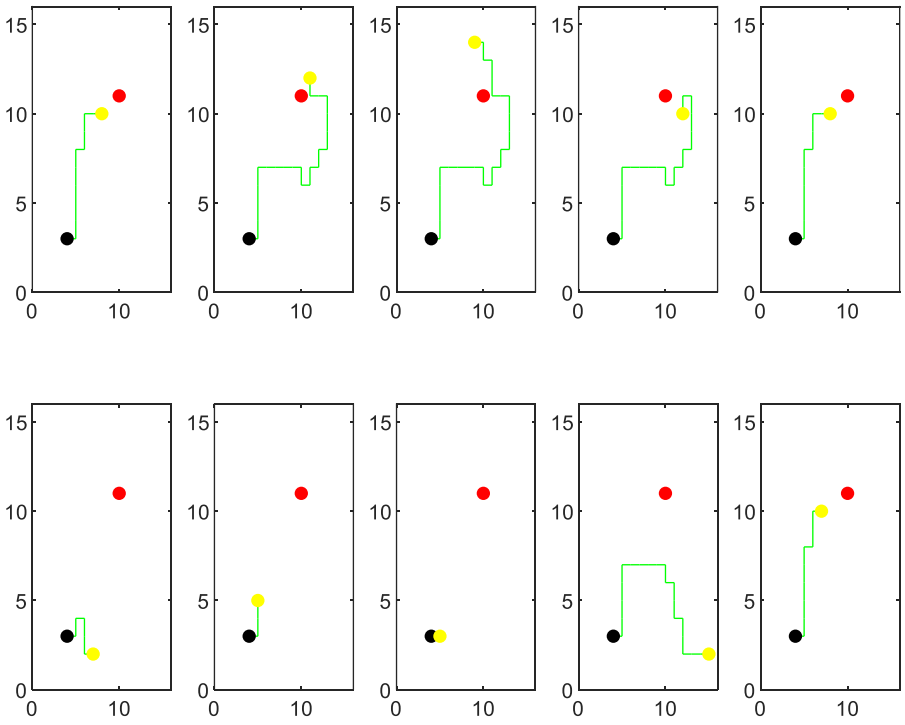

upward movement
10
5
0
-5
-10
15
10
5
0
0
5
10
15


horizontal movement to the right
4
2
0
-2
-4
-6
-8
15
10
5
0
0
5
10
15



3. In this part we try to find the effect of learning rate and discount factor on the learning process.
   - Learning rate: learning rate is correlated with the speed of learning. We may lose some optimums conditions by increasing this amount but we will reach one of them quicker.

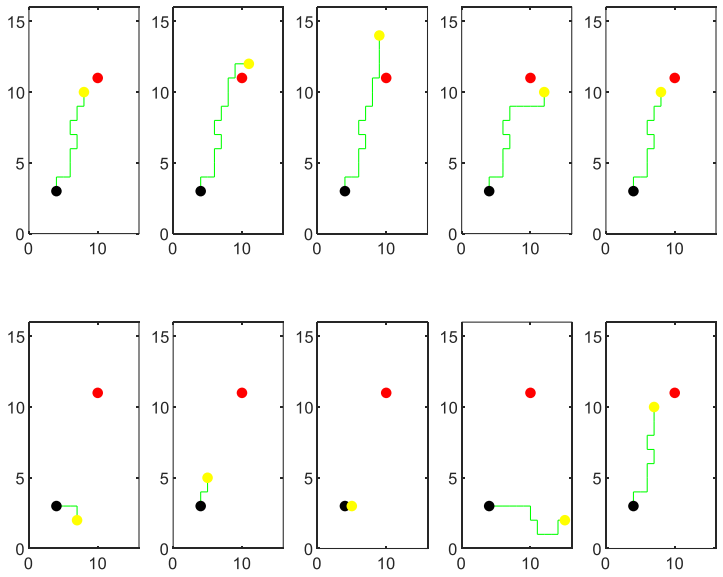
Learning rate=-0.5 , N=10000



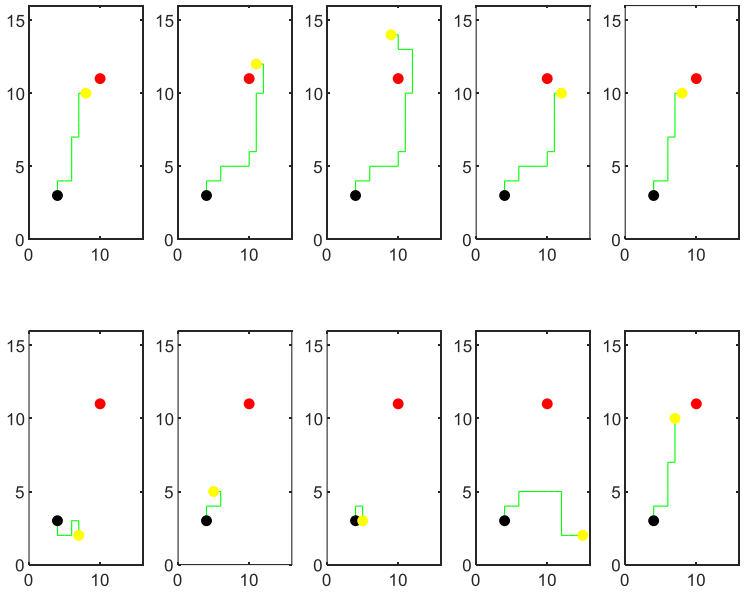
Advanced Topics in Neuroscience
Instructor: Dr. Ali Ghazizadeh

problem set #6 (reinforcement learning)
Name: Sana Aminnaji
Student ID: 98104722

Learning rate=0.8 , N=7000

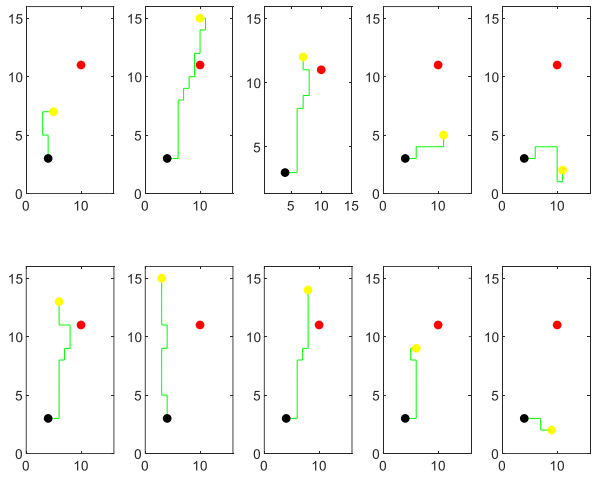


Learning rate=0.9 , N=7000

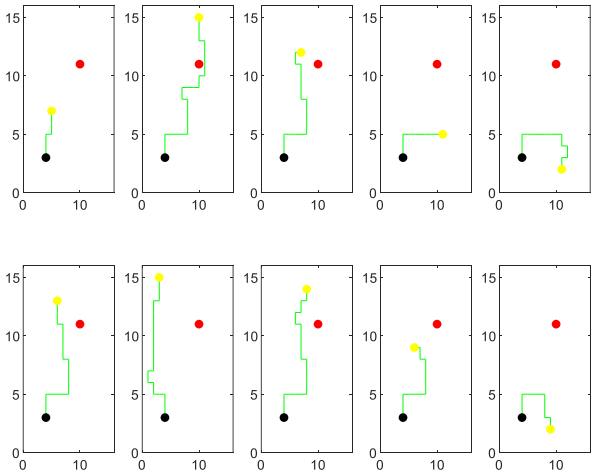


- Discount factor: The discount factor adjusts the importance of rewards over time. The later we receive rewards, the less attractive they are to present calculations.

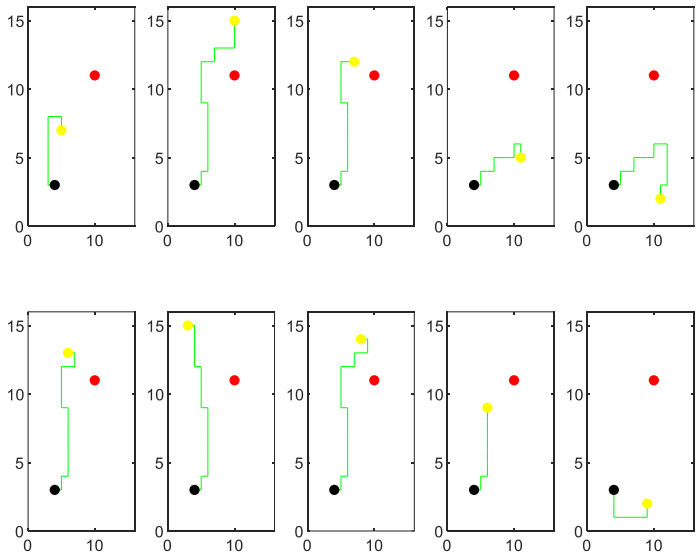
Discount factor = 0.2



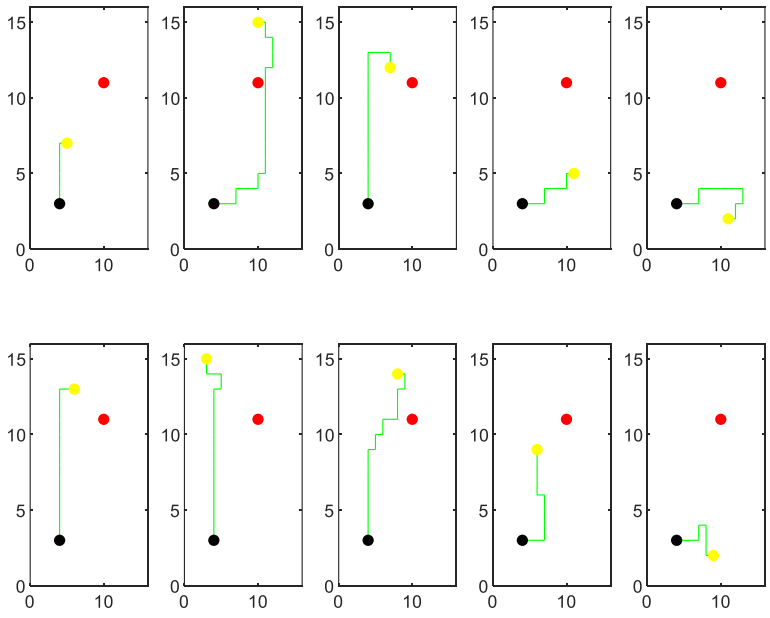
Discount factor = 0.3



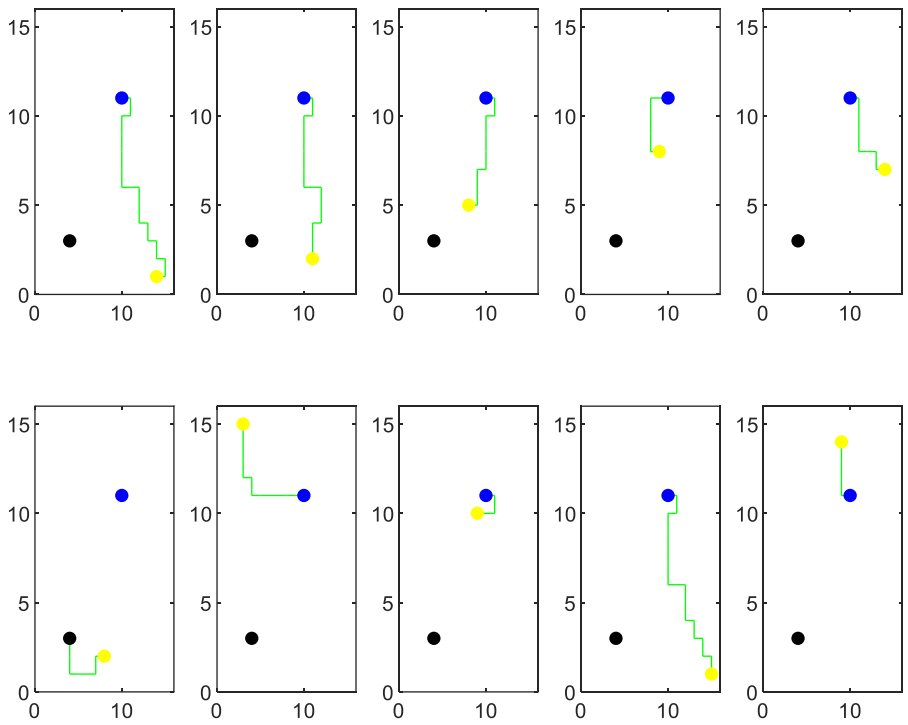
Discount factor = 0.5



Discount factor = 0.8



4. In this section we will have two different targets and watch what will happen:
   In the figure below, the weight of blue target is twice the weight of black target:



5. We tried to learn the agent using Q-learning approach, we can have different policies for learning.
   The estimated error in this approach is calculated using TD algorithm.
   Changing the policy will create exploring the environment more or in the other cases finding the goal with more speed.