Implement a Basic Driving Agent

QUESTION: *Observe what you see with the agent's behavior as it takes random actions. Does the* ***smartcab*** *eventually make it to the destination? Are there any other interesting observations to note?*

Answer: Sometimes smartcab reaches destination. Since the smartcab is not trained it always takes random action and get poor rewards as it violates laws and also misses the destination based on random actions.

Inform the Driving Agent

QUESTION: *What states have you identified that are appropriate for modeling the* ***smartcab*** *and environment? Why do you believe each of these states to be appropriate for this problem?*

Answer:The states that are appropriate for modeling are

inputs['light'], inputs['oncoming'], inputs['left'],inputs['right'],self.next_waypoint

The reason why I choose these states are the inputs related to light, oncoming traffic, left and right will make sure the agent follows the traffic rules and punish them if the agent takes an action to violate these inputs. Once the agent explores all the states the agent can drive based on the input states it sees and follow the traffic rules. The self.next_waypoint is the direction recommended to reach the destination, the reason why I choose this is if there is a green light with no traffic but if destination is forward but if agent makes left it will get a negative reward and if the agent goes forward it gets maximum reward, so when the agent encounters the same state then it will decide based on the next_waypoint to reach the detination otherwise agent might or might not reach destination on time even though it is following traffic rules.

OPTIONAL: *How many states in total exist for the* ***smartcab*** *in this environment? Does this number seem reasonable given that the goal of Q-Learning is to learn and make informed decisions about each state? Why or why not?*

Answer: There are total 51200 states for the smartcab in this environment. 2 (lights) x 4 (oncoming) x 4 (left) * 4 (Right) * 4 (next_waypoint) * 100 (max deadline). I choose 5 input only and ignored deadline as the deadline for example of 100 adds 100 values for deadline and adds more combinations with different inputs and the next_way_point and agent might not be able to learn everything from the iterations we have and might take more time to explore. The 5 inputs look reasonable as agent need to learn each state and what action to take at each state as these are all the possible states that reflect traffic rules and path to reach destination.

Implement a Q-Learning Driving Agent

QUESTION: *What changes do you notice in the agent's behavior when compared to the basic driving agent when random actions were always taken? Why is this behavior occurring?*

Answer: When the agent is taking actions randomly it violates the traffic rules and also it misses the destination even though the destination is close by because of random actions. Once we implement the Q learning driving agent based on the Q values (some combination of instant reward and future reward) for the state, action combinations it will choose the best action with max q value for a particular state. Once the agent learns a state and different actions for that state, it will try to follow the max reward action which basically means following traffic rules and trying to reach destination using self.next_waypoint. This improves agent reaching destination most of the time compared to random actions.

Improve the Q-Learning Driving Agent

QUESTION: *Report the different values for the parameters tuned in your basic implementation of Q-Learning. For which set of parameters does the agent perform best? How well does the final driving agent perform?*

Answer. Below are the different values for the parameters tuned in basic implementation of Q Learning. The agent performed well for Alpha 0.5 and gamma 0.1. With Alpha 0.5 and gamma 0.1, the agent was able to reach the destination 100% of the time in last 10 trails. In overall 100 trails it reached an average of 98.5 times the destination.

ALPHA	GAMMA	Total Negative rewards in all trails	Sum of negative reward in last 10 trails	Total steps in all trails	Sum of steps in last 10 trails	Reached destination in all 100 trails	Reached destination in last 10 trails
0.1	0.1	-32	-0.5	1418	119	98	9
0.1	0.1	-37	-1	1361	128	100	10
0.2	0.1	-42	-1.5	1426	115	100	10
0.2	0.1	-49	-3	1369	133	98	10
0.2	0.2	-54	-2.5	1327	138	98	10
0.2	0.2	-51	-0.5	1521	124	97	10
0.5	0.1	-52	-0.5	1273	123	99	10
0.5	0.1	-55	-0.5	1363	150	98	10
0.5	0.5	-71	-9.5	1378	138	98	10
0.5	0.5	-118	-4	1513	160	92	9
0.9	0.5	-82	-0.5	1629	149	91	10
0.9	0.5	-66	-1.5	1410	107	97	10
0.9	0.7	-114	-1	1624	128	82	10
0.9	0.7	-50	-6	1356	142	99	10

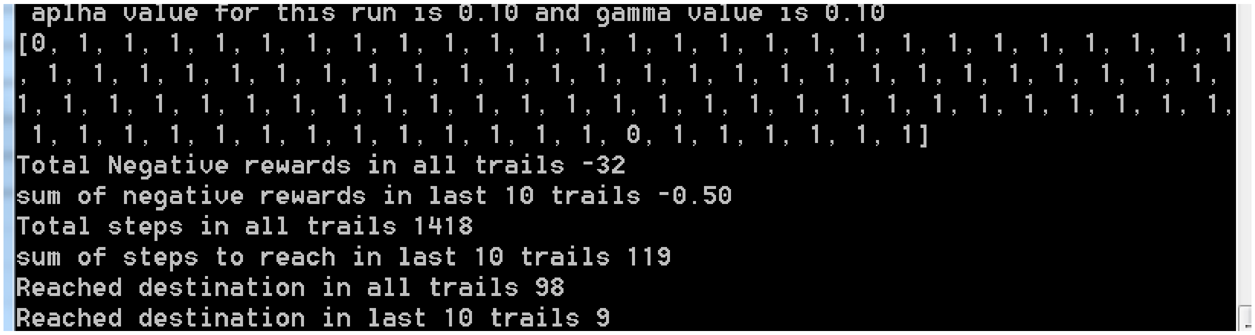
QUESTION: *Does your agent get close to finding an optimal policy, i.e. reach the destination in the minimum possible time, and not incur any penalties? How would you describe an optimal policy for this problem?*

Answer: With alpha 0.5 and gamma 1 , I think the agent got close to finding an optimal policy as it is reaching the destination almost all the time once it explored the states as it can been seen that it reached destination 100% in last 10 trails. Also the negative rewards it encountered once it explored the states is less as total negative rewards it encountered in last 10 trails is of average -0.5. The number of steps it took to reach the destination is also not high from below table as it reached destination with an average of 13.65 steps ((123+150/2)/10) per trail.

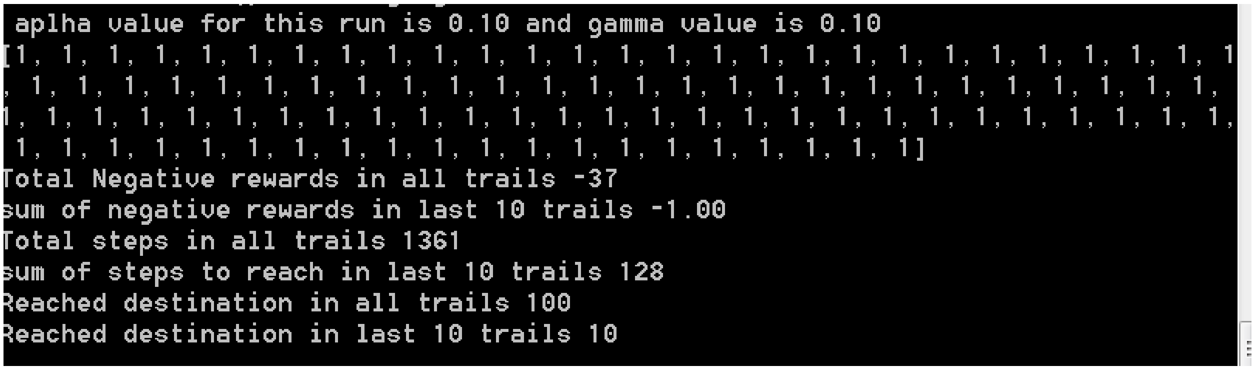
ALPHA	GAMMA	Total Negative rewards in all trails	Sum of negative reward in last 10 trails	Total steps in all trails	Sum of steps in last 10 trails	Reached destination in all 100 trails	Reached destination in last 10 trails
0.1	0.1	-32	-0.5	1418	119	98	9
0.1	0.1	-37	-1	1361	128	100	10
0.2	0.1	-42	-1.5	1426	115	100	10
0.2	0.1	-49	-3	1369	133	98	10
0.2	0.2	-54	-2.5	1327	138	98	10
0.2	0.2	-51	-0.5	1521	124	97	10
0.5	0.1	-52	-0.5	1273	123	99	10
0.5	0.1	-55	-0.5	1363	150	98	10
0.5	0.5	-71	-9.5	1378	138	98	10
0.5	0.5	-118	-4	1513	160	92	9
0.9	0.5	-82	-0.5	1629	149	91	10
0.9	0.5	-66	-1.5	1410	107	97	10
0.9	0.7	-114	-1	1624	128	82	10
0.9	0.7	-50	-6	1356	142	99	10

Below are the screenshots based on which the above table is constructed.

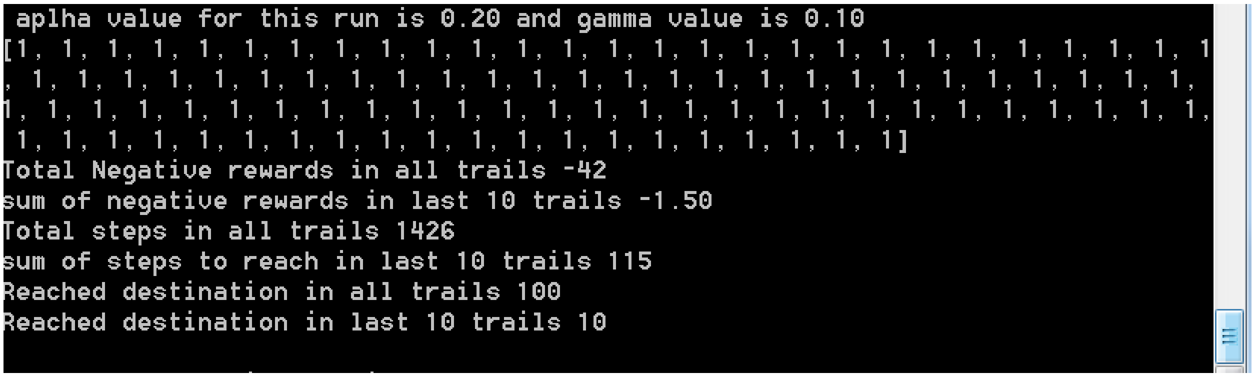
In the below screenshots 1 means it reached destination 0 means it didn’t reach destination



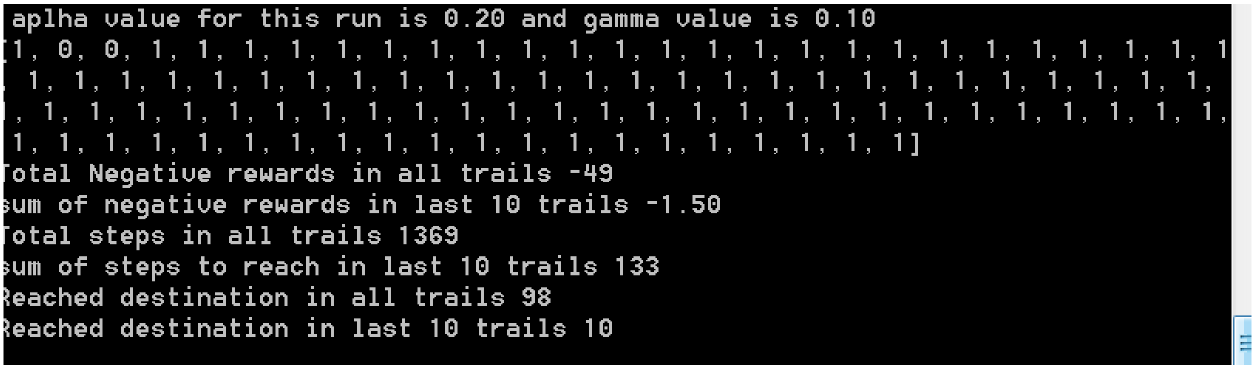
```
 aplha value for this run is 0.10 and gamma value is 0.10
[0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1]
Total Negative rewards in all trails -32
sum of negative rewards in last 10 trails -0.50
Total steps in all trails 1418
sum of steps to reach in last 10 trails 119
Reached destination in all trails 98
Reached destination in last 10 trails 9
```



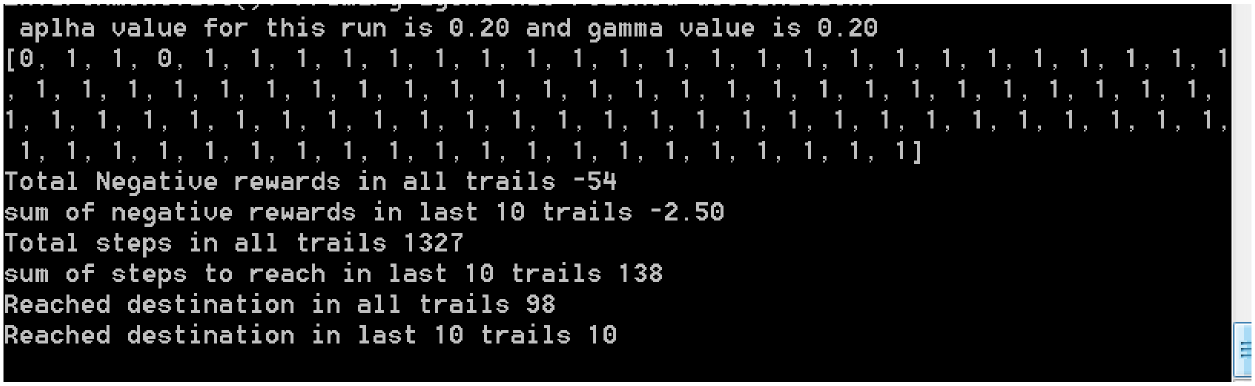
```
 aplha value for this run is 0.10 and gamma value is 0.10
[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]
Total Negative rewards in all trails -37
sum of negative rewards in last 10 trails -1.00
Total steps in all trails 1361
sum of steps to reach in last 10 trails 128
Reached destination in all trails 100
Reached destination in last 10 trails 10
```



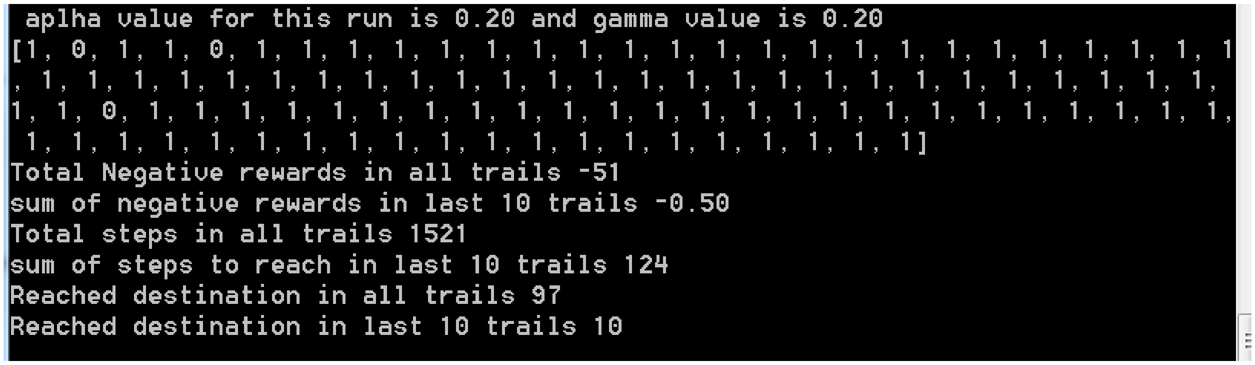
```
 aplha value for this run is 0.20 and gamma value is 0.10
[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]
Total Negative rewards in all trails -42
sum of negative rewards in last 10 trails -1.50
Total steps in all trails 1426
sum of steps to reach in last 10 trails 115
Reached destination in all trails 100
Reached destination in last 10 trails 10
```



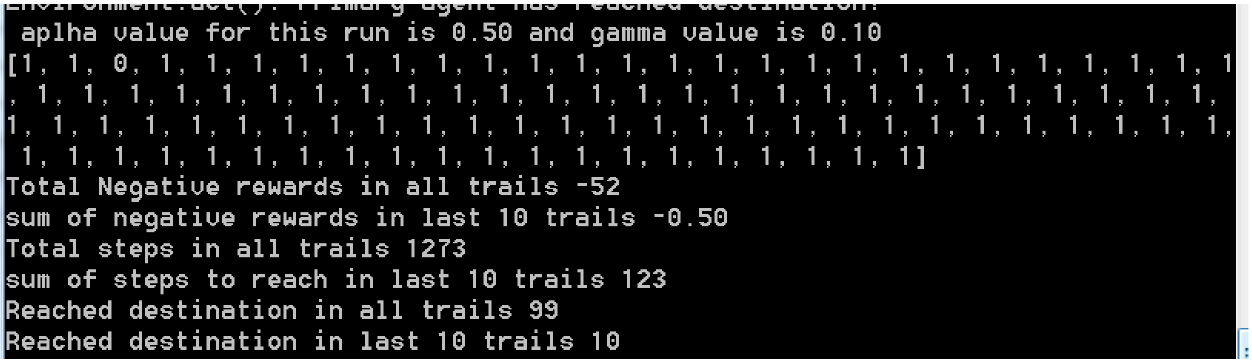
```
 aplha value for this run is 0.20 and gamma value is 0.10
[1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]
Total Negative rewards in all trails -49
sum of negative rewards in last 10 trails -1.50
Total steps in all trails 1369
sum of steps to reach in last 10 trails 133
Reached destination in all trails 98
Reached destination in last 10 trails 10
```



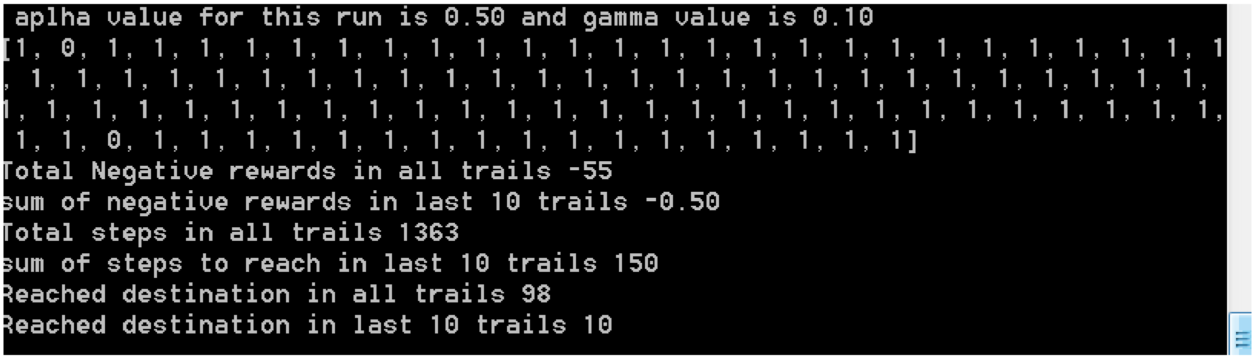
```
 aplha value for this run is 0.20 and gamma value is 0.20
[0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]
Total Negative rewards in all trails -54
sum of negative rewards in last 10 trails -2.50
Total steps in all trails 1327
sum of steps to reach in last 10 trails 138
Reached destination in all trails 98
Reached destination in last 10 trails 10
```



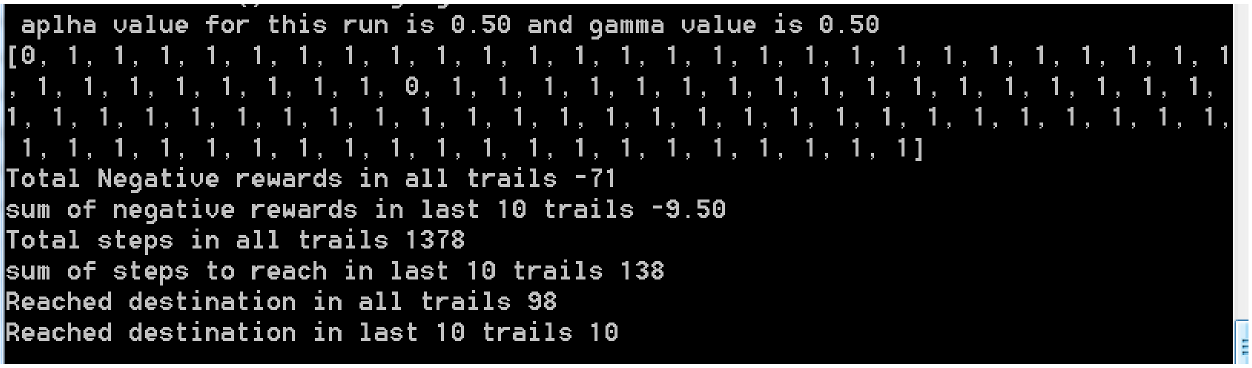
```
 aplha value for this run is 0.20 and gamma value is 0.20
[1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]
Total Negative rewards in all trails -51
sum of negative rewards in last 10 trails -0.50
Total steps in all trails 1521
sum of steps to reach in last 10 trails 124
Reached destination in all trails 97
Reached destination in last 10 trails 10
```



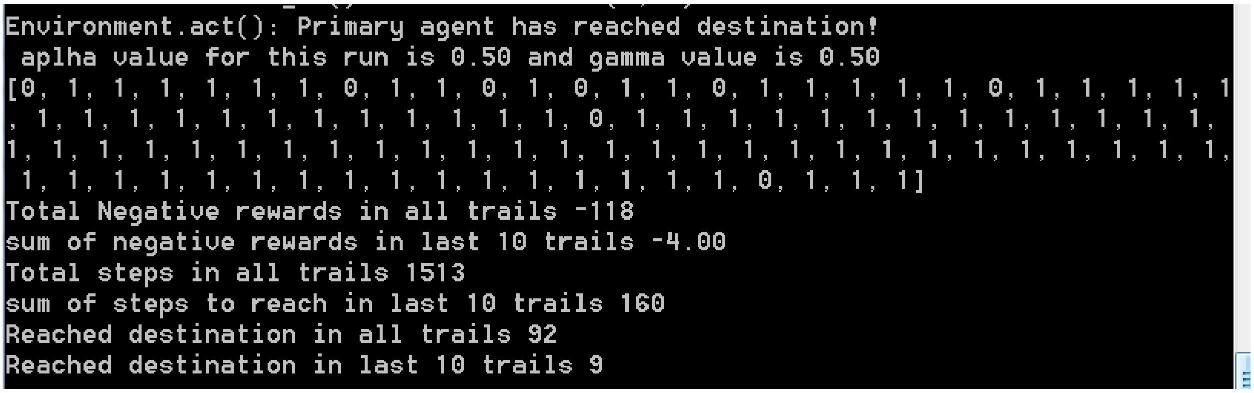
```
Environment.act(): Primary agent has reached destination!
 aplha value for this run is 0.50 and gamma value is 0.10
[1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]
Total Negative rewards in all trails -52
sum of negative rewards in last 10 trails -0.50
Total steps in all trails 1273
sum of steps to reach in last 10 trails 123
Reached destination in all trails 99
Reached destination in last 10 trails 10
```



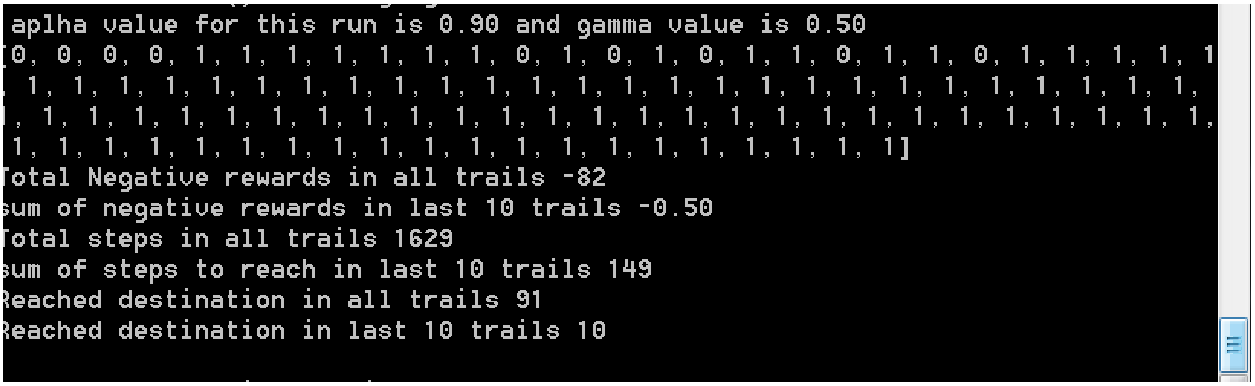
```
 aplha value for this run is 0.50 and gamma value is 0.10
[1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]
Total Negative rewards in all trails -55
sum of negative rewards in last 10 trails -0.50
Total steps in all trails 1363
sum of steps to reach in last 10 trails 150
Reached destination in all trails 98
Reached destination in last 10 trails 10
```



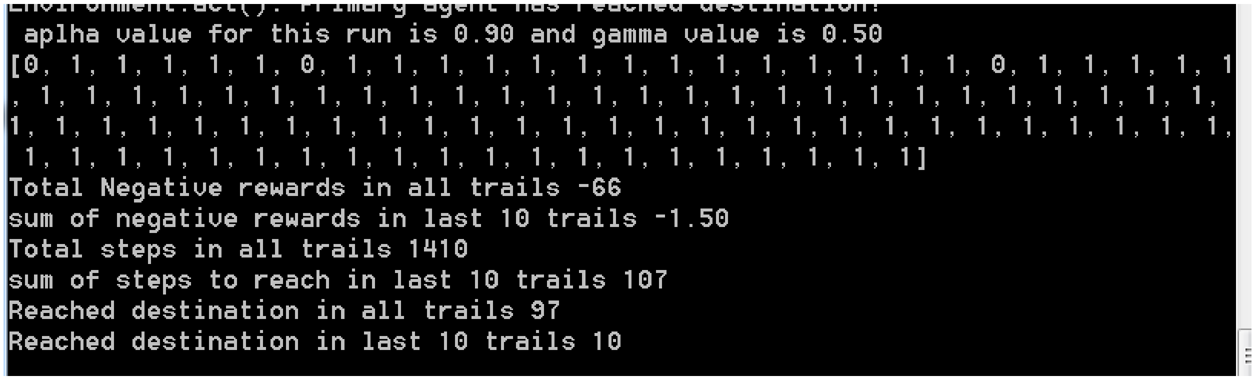
```
 aplha value for this run is 0.50 and gamma value is 0.50
[0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]
Total Negative rewards in all trails -71
sum of negative rewards in last 10 trails -9.50
Total steps in all trails 1378
sum of steps to reach in last 10 trails 138
Reached destination in all trails 98
Reached destination in last 10 trails 10
```



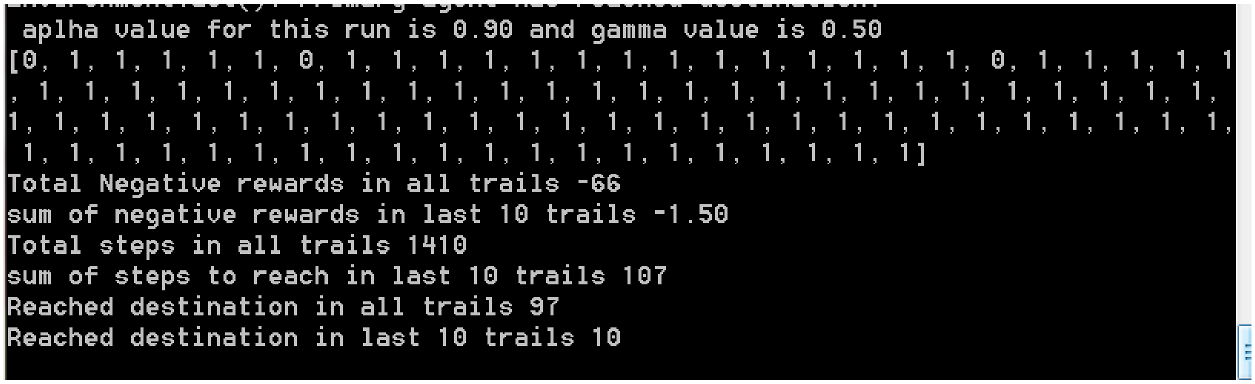
```
Environment.act(): Primary agent has reached destination!
 aplha value for this run is 0.50 and gamma value is 0.50
[0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1
, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1]
Total Negative rewards in all trails -118
sum of negative rewards in last 10 trails -4.00
Total steps in all trails 1513
sum of steps to reach in last 10 trails 160
Reached destination in all trails 92
Reached destination in last 10 trails 9
```



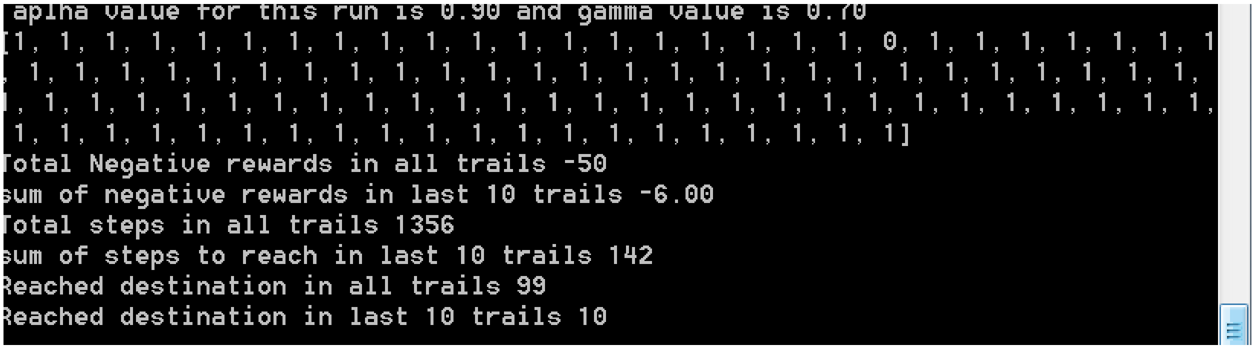
```
 aplha value for this run is 0.90 and gamma value is 0.50
[0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1
, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]
Total Negative rewards in all trails -82
sum of negative rewards in last 10 trails -0.50
Total steps in all trails 1629
sum of steps to reach in last 10 trails 149
Reached destination in all trails 91
Reached destination in last 10 trails 10
```



```
Environment.act(): Primary agent has reached destination!
 aplha value for this run is 0.90 and gamma value is 0.50
[0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1
, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]
Total Negative rewards in all trails -66
sum of negative rewards in last 10 trails -1.50
Total steps in all trails 1410
sum of steps to reach in last 10 trails 107
Reached destination in all trails 97
Reached destination in last 10 trails 10
```



```
 aplha value for this run is 0.90 and gamma value is 0.50
[0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1
, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]
Total Negative rewards in all trails -66
sum of negative rewards in last 10 trails -1.50
Total steps in all trails 1410
sum of steps to reach in last 10 trails 107
Reached destination in all trails 97
Reached destination in last 10 trails 10
```



```
 aplha value for this run is 0.90 and gamma value is 0.70
[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1
, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]
Total Negative rewards in all trails -50
sum of negative rewards in last 10 trails -6.00
Total steps in all trails 1356
sum of steps to reach in last 10 trails 142
Reached destination in all trails 99
Reached destination in last 10 trails 10
```