Assignment-03 Q-Learning

Student	Piyush Kumar
SrNo.	23801
Course	UMC 203: Artificial Intelligence and Machine Learning
Date	8 April 2025

Q-Learning

In this assignment, you will implement Q-Learning to find the route in a grid cell-based environment. Your agent has to find its way from the start cell to the goal cell in a grid maze. Along the way there are special cells called traps and boosts which give your agent negative and positive rewards respectively. Youare provided with three files, **BFS.ipynb, QL Assignment.ipynb** and **themes.json. BFS.ipynb** performs a Breadth-first search on a grid maze to find if a path exists. QL **Assignment.ipynb** has the environment and agent related code for training and test ing. It generates a **pickle file (.pkl)** for each agent training scenario.

Q-Learning is a model-free reinforcement learning algorithm that learns the value of an action in a particular state. It does this by using the Bellman equation to update the Q-values based onthe rewards received from taking actions in the environment. The Q-value for a state-action pair is updated using the formula:

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left(r + \gamma \max_{a'} Q(s',a') - Q(s,a) \right)$$

where:

- $Q(s,a)$ is the Q-value for state s and action a.
- α is the learning rate ($0 < \alpha < 1$).
- r is the reward received after taking action a in state s.
- γ is the discount factor ($0 < \gamma < 1$).
- s' is the next state after taking action a.
- $\max_{a'} Q(s',a')$ is the maximum Q-value for the next state s' over all possible actions a'.
- $Q(s,a)$ is the current Q-value for state s and action a.

The Q-learning algorithm works by iteratively updating the Q-values for each state-action pair based on the rewards received from the environment. The algorithm continues to learn until the Q-values converge to their optimal values.

Question

(a). Complete the QL Assignment.ipynb file. Regions where you have to fill in your code have been marked. Use your SR.No to generate a maze unique to you. Use the BFS.ipynb file to ensure that a path exists in your generated maze (we have made sure that a path exists for all your SR.No but it's better to check just to stay on the safer side). (10 marks)

File has been attached with the Assignment.

(b). Train your agent on two scenarios, one where traps and boots are disabled and another where they are enabled, and comment on the paths learned by the agents in these scenarios. If the number of steps taken by your agent, when traps and boosts are disabled, are same as the number of steps in the path generated by **BFS.ipynb** then you are on the right track. (2 marks)

Hyperparameter	Value
REWARD_GOAL	100
REWARD_TRAP	-1000
REWARD_OBSTACLE	-1000
REWARD_REVISIT	-10
REWARD_ENEMY	-2000
REWARD_STEP	-1
REWARD_BOOST	20
GAMMA	0.99
ALPHA	0.8
EPSILON	1.0
EPSILON_DECAY	0.9995
N_EPISODES	10000
EPSILON_MIN	0.1
MAX_STEPS	300

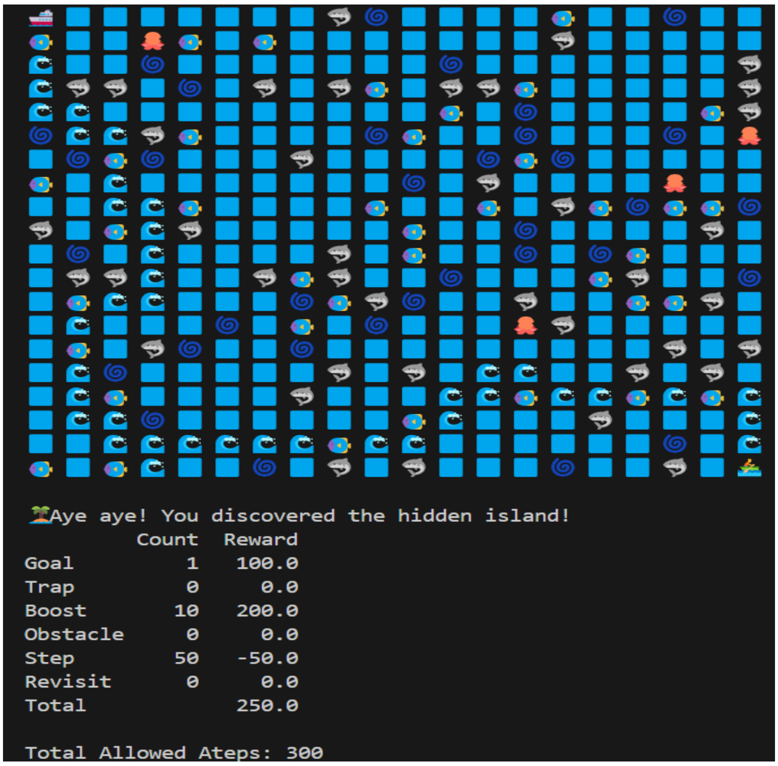
Table 1: Hyperparameters used in the Q-learning algorithm.

Trap and Boost Disabled:

	Count	Reward
Goal	1	100.0
Boost	10	200.0
Step	50	-50.0
Total		250.0

Table 2: Path taken by the agent when traps and boots are disabled (Sample 1).

.



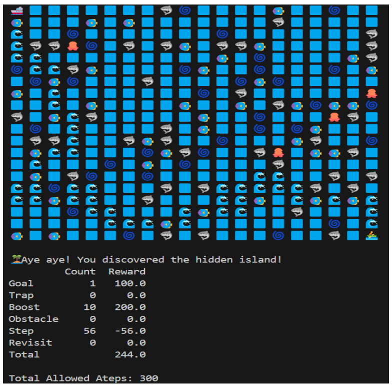
```
🏝Aye aye! You discovered the hidden island!
              Count   Reward
Goal              1    100.0
Trap              0      0.0
Boost            10    200.0
Obstacle          0      0.0
Step             50    -50.0
Revisit           0      0.0
Total                  250.0

Total Allowed Ateps: 300
```

PKL File has been attached with the Assignment.

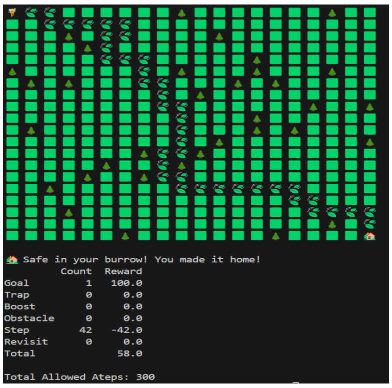
{PKL File has been attached with the Assignment.

	Count	**Reward**
Goal	1	100.0
boost	10	200.0
Step	50	-50.0
Total		250.0

Table 3: Path taken by the agent when traps and boots are disabled (Sample 2).



```
🏝Aye aye! You discovered the hidden island!
          Count  Reward
Goal          1   100.0
Trap          0     0.0
Boost        10   200.0
Obstacle      0     0.0
Step         56   -56.0
Revisit       0     0.0
Total             244.0

Total Allowed Ateps: 300
```



```
🏠 Safe in your burrow! You made it home!
          Count  Reward
Goal          1   100.0
Trap          0     0.0
Boost         0     0.0
Obstacle      0     0.0
Step         42   -42.0
Revisit       0     0.0
Total              58.0

Total Allowed Ateps: 300
```

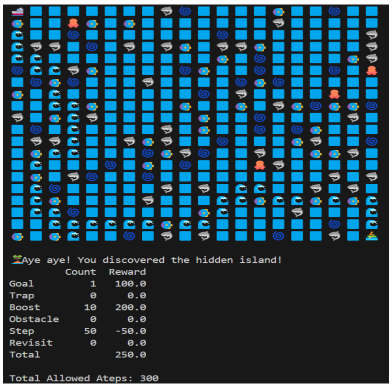
Figure1- Enable, **Figure2- Disabled**

(c). Scenario 1:

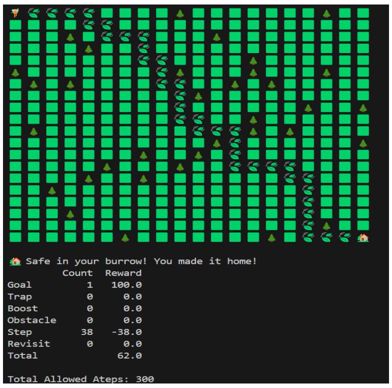
In my first scenario I have changed the REWARD_STEP where I increased the reward from the −1 to 60 such that now it is taking more steps as compared to the previous one, where first REWARD_STEP was −1.

Hyperparameter	**Value**
REWARD_GOAL	100
REWARD_TRAP	-1000
REWARD_OBSTACLE	-1000
REWARD_REVISIT	-10
REWARD_ENEMY	-2000
REWARD_STEP	-1
REWARD_BOOST	20

Table 4: Hyperparameters used in the Q-learning algorithm for the above Data.



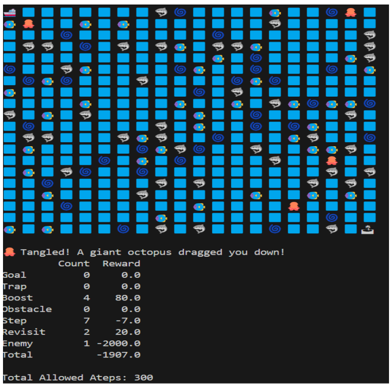
```
🏝Aye aye! You discovered the hidden island!
          Count  Reward
Goal          1   100.0
Trap          0     0.0
Boost        10   200.0
Obstacle      0     0.0
Step         50   -50.0
Revisit       0     0.0
Total             250.0

Total Allowed Ateps: 300
```



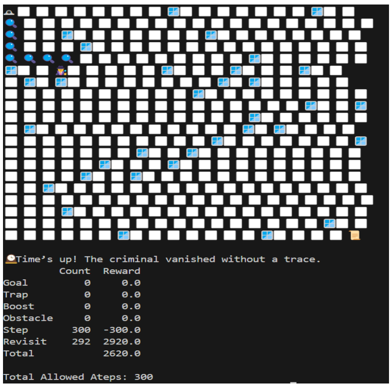
```
🏠 Safe in your burrow! You made it home!
          Count  Reward
Goal          1   100.0
Trap          0     0.0
Boost         0     0.0
Obstacle      0     0.0
Step         38   -38.0
Revisit       0     0.0
Total              62.0

Total Allowed Ateps: 300
```

Figure1- Enable, **Figure2- Disabled**

In another scenario i just changed the REWARD REVISIT variable to 10 such that now it is not even moving to the goal. The agent is just revisiting the same state again and again.



```
Tangled! A giant octopus dragged you down!
          Count   Reward
Goal          0      0.0
Trap          0      0.0
Boost         4     80.0
Obstacle      0      0.0
Step          7     -7.0
Revisit       2     20.0
Enemy         1  -2000.0
Total            -1907.0

Total Allowed Ateps: 300
```



```
Time’s up! The criminal vanished without a trace.
          Count   Reward
Goal          0      0.0
Trap          0      0.0
Boost         0      0.0
Obstacle      0      0.0
Step        300   -300.0
Revisit     292   2920.0
Total             2620.0

Total Allowed Ateps: 300
```

Figure1- Enable,**Figure2- Disabled**

In my case the number of steps taken by the agent when traps and boots are disabled is 56 and 50, respectively. But, the number of steps taken by the Path.ipynb are 38.

.

(d).**Steps 1**:

$$Q((0,\ 0),\ 2) \leftarrow 0.00 + 0.1\,[-2000 + 0.99 \cdot 0.00 - 0.00] = -200.0000$$

Steps 2 :

$$Q((0,\ 0),\ 0) \leftarrow 0.00 + 0.1\,[-2100 + 0.99 \cdot 0.00 - 0.00] = -210.0000$$

Steps 3 :

$$Q((0,\ 0),\ 3) \leftarrow 0.00 + 0.1\,[-1 + 0.99 \cdot 0.00 - 0.00] = -0.1000$$

Steps 4 :

$$Q((0,\ 1),\ 1) \leftarrow 0.00 + 0.1\,[-1 + 0.99 \cdot 0.00 - 0.00] = -0.1000$$

Steps 5 :

$$Q((1,\ 1),\ 3) \leftarrow 0.00 + 0.1\,[-1 + 0.99 \cdot 0.00 - 0.00] \;=\; -0.1000$$

NOTE This pickle pickle for this is the 23801 Disabled 1.pkl.