CS156 (Introduction to AI), Spring 2022

Homework 11 submission

Roster Name: Bernard Tan

Preferred Name (if different): Bernard

Student ID: 015215317

Email address: bernard.tan@sjsu.edu

Any special notes or anything you would like to communicate to me about this homework submission goes in here.

References and sources

List all your references and sources here. This includes all sites/discussion boards/blogs/posts/etc. where you grabbed some code examples.

Solution

Load libraries and set random number generator seed

```
import numpy as np
import pandas as pd
import gym
```

```
np.random.seed(42)
```

```
env = gym.make("FrozenLake-v0", is_slippery=False).env
```

```
env.seed(42)
env.reset()
env.render()
```

```
print("Action Space {}".format(env.action_space))
print("State Space {}".format(env.observation_space))
```

```
SFFF
FHFH
FFFH
HFFG
Action Space Discrete(4)
State Space Discrete(16)
```

```
qtable = np.zeros([env.observation_space.n, env.action_space.n]) #You could also make this dy
discount = 0.9
learningrate = 0.9
epsilon = 0.2
for episode in range(1,10000):
    done = False
    reward_total = 0
    state = env.reset()
    while done != True:
        explore_eploit = np.random.uniform(0, 1)
        if explore_eploit < epsilon:
            action = env.action_space.sample() # explore action space
        else:
            action = np.argmax(qtable[state]) # exploit learned values

        state_new, reward, done, info = env.step(action) #take the action
        qtable[state,action] += learningrate * (reward + discount * np.max(qtable[state_new,:
        reward_total = reward_total + reward
        state = state_new
    if episode % 50 == 0:
        print('Episode {} Total Reward: {}'.format(episode,reward_total))
```

```
Episode 7000 Total Reward: 0.0
Episode 7050 Total Reward: 0.0
Episode 7100 Total Reward: 0.0
Episode 7150 Total Reward: 0.0
Episode 7200 Total Reward: 0.0
Episode 7250 Total Reward: 0.0
Episode 7300 Total Reward: 0.0
Episode 7350 Total Reward: 0.0
Episode 7400 Total Reward: 0.0
Episode 7450 Total Reward: 0.0
Episode 7500 Total Reward: 0.0
Episode 7550 Total Reward: 0.0
Episode 7600 Total Reward: 0.0
Episode 7650 Total Reward: 0.0
Episode 7700 Total Reward: 0.0
Episode 7750 Total Reward: 0.0
Episode 7800 Total Reward: 0.0
Episode 7850 Total Reward: 0.0
Episode 7900 Total Reward: 0.0
Episode 7950 Total Reward: 0.0
```

```
Episode 7950 Total Reward: 0.0
Episode 8000 Total Reward: 0.0
Episode 8050 Total Reward: 0.0
Episode 8100 Total Reward: 0.0
Episode 8150 Total Reward: 0.0
Episode 8200 Total Reward: 1.0
Episode 8250 Total Reward: 0.0
Episode 8300 Total Reward: 1.0
Episode 8350 Total Reward: 1.0
Episode 8400 Total Reward: 1.0
Episode 8450 Total Reward: 1.0
Episode 8500 Total Reward: 0.0
Episode 8550 Total Reward: 1.0
Episode 8600 Total Reward: 1.0
Episode 8650 Total Reward: 1.0
Episode 8700 Total Reward: 1.0
Episode 8750 Total Reward: 1.0
Episode 8800 Total Reward: 1.0
Episode 8850 Total Reward: 1.0
Episode 8900 Total Reward: 1.0
Episode 8950 Total Reward: 1.0
Episode 9000 Total Reward: 1.0
Episode 9050 Total Reward: 1.0
Episode 9100 Total Reward: 1.0
Episode 9150 Total Reward: 1.0
Episode 9200 Total Reward: 1.0
Episode 9250 Total Reward: 0.0
Episode 9300 Total Reward: 0.0
Episode 9350 Total Reward: 1.0
Episode 9400 Total Reward: 1.0
Episode 9450 Total Reward: 0.0
Episode 9500 Total Reward: 0.0
Episode 9550 Total Reward: 0.0
Episode 9600 Total Reward: 1.0
Episode 9650 Total Reward: 1.0
Episode 9700 Total Reward: 1.0
Episode 9750 Total Reward: 1.0
Episode 9800 Total Reward: 1.0
Episode 9850 Total Reward: 1.0
Episode 9900 Total Reward: 0.0
```

```
print(qtable)
```

```
[[0.531441   0.59049    0.4782969  0.531441  ]
 [0.531441   0.         0.43046721 0.47829685]
 [0.4782969  0.         0.         0.        ]
 [0.         0.         0.         0.        ]
 [0.59049    0.6561     0.         0.531441  ]
 [0.         0.         0.         0.        ]
 [0.         0.         0.         0.        ]
 [0.         0.         0.         0.        ]
 [0.6561     0.         0.729      0.59049   ]
 [0.6561     0.81       0.81       0.        ]
 [0.729      0.9        0.         0.        ]
 [0.         0.         0.         0.        ]
 [0.         0.         0.         0.        ]]
```

```
       [0.         0.81       0.9        0.729     ]
       [0.81       0.9        1.         0.81      ]
       [0.         0.         0.         0.        ]]
```

```
# Let's see how the algorithm solves the taxi game by following the policy to take actions de

reward_total=0
obs= env.reset()
env.render()

done=False

while done != True:
    action = np.argmax(qtable[obs])
    obs, reward, done, info = env.step(action) #take step using selected action
    reward_total = reward_total + reward
    env.render()
#Print the reward of these actions
print("Total reward is %r" % reward_total)
```

```
    SFFF
    FHFH
    FFFH
    HFFG
      (Down)
    SFFF
    FHFH
    FFFH
    HFFG
      (Down)
    SFFF
    FHFH
    FFFH
    HFFG
      (Right)
    SFFF
    FHFH
    FFFH
    HFFG
      (Down)
    SFFF
    FHFH
    FFFH
    HFFG
      (Right)
    SFFF
    FHFH
    FFFH
    HFFG
      (Right)
    SFFF
    FHFH
    FFFH
```

```
HFFG
Total reward is 1.0
```

✓ 0s completed at 11:35 AM