CMPE 58Y - Robot Learning
Homework 1: Q-learning

February 27, 2020

1 Introduction

In this homework you will implement Q-learning for the cart pole task [1] in OpenAI Gym environment which is written for `python`. You can find the instructions for installation at OpenAI Gym's website. The following is a quick example to the environment:

```
import gym
env = gym.make("CartPole-v1")
env.reset()
for _ in range(1000):
    # you don't have to render. it's just for visualization.
    env.render()
    # take a random action
    observation, reward, done, info = env.step(env.action_space.sample())
env.close()
```

In this script, we take random actions. After doing an action, the environment provides you a new `observation` and `reward`. `done` is a boolean which denotes whether the trajectory should be terminated or not. This is generally used for constructing a loop such as: `while not done: do stuff`. However we will not use this variable as it makes things quite easy for this task. Instead, terminate the trajectory after 500 timesteps. You can consider the task is solved if you consecutively get rewards higher than 400.

The convergence of the algorithm depends on your hyperparameter settings. Choose them wisely. One of the most important hyperparameters in this problem is quantizing continuous states. Use the following method and variables to understand your environment:

```
env.observation_space.sample()
env.observation_space.low
env.observation_space.high
```

2 Deliverables

Plot the reward over episodes. Submit your code (a jupyter notebook is also fine) to `ahmetoglu.alper@gmail.com`. For any questions regarding the description, environment installation, hyperparameters and so on, you can come to my office BM-31 (COLORS-LAB). Cheating will be penalized by -200 points.

Deadline: Tuesday, 3 March, 9:00 A.M. (You will be graded out of 100)
Late deadline: Sunday, 8 March, 11:59 P.M. (You will be graded out of 80)

References

[1] Andrew G Barto, Richard S Sutton, and Charles W Anderson. Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE transactions on systems, man, and cybernetics*, (5):834–846, 1983.