Assignment 3

叶增渝 123033910090

一、Cliff Enviroment Set Up:

文件使用 ipynb 形式进行组织

（1）第一个 cell 导入了必要的包;

（2）第二个 cell 设定了基础参数, 包括 Cliff world size、α、ε、γ、start position、end position、training epochs 等

（3）第三个 cell 包含了进行 Sarsa 与 Q-learning 时需要用到的一系列子函数与展示 policy 路线的 show_policy()函数

（4）第四个 cell 包含了进行 epoch learning 的 Sarsa 与 Q-learning 方法

（5）第五个 cell 进行了 2 种方法训练与展示

二、ε exploration:

我们设定 epoch 为 10000，地图大小为（5，10），α为 0.2，γ为 0.5

当ε为 0.1 时，其对 Sarsa 方法的影响较大，往往使得 Sarsa 无法找到最优路线，但对 Q-learning 方法无影响，以下为重复 5 次的结果（去除重复）:

Sarsa:

(4, 0)->(3, 0)->(2, 0)->(1, 0)->(0, 0)->(0, 1)->(0, 2)->(0, 3)->(0, 4)->(0, 5)->(0, 6)->(0, 7)->(0, 8)->(0, 9)->(1, 9)->(2, 9)->(3, 9)->(4, 9)

(4, 0)->(3, 0)->(2, 0)->(2, 1)->(1, 1)->(0, 1)->(0, 2)->(0, 3)->(0, 4)->(0, 5)->(0, 6)->(0, 7)->(0, 8)->(0, 9)->(1, 9)->(2, 9)->(3, 9)->(4, 9)

(4, 0)->(3, 0)->(2, 0)->(2, 1)->(2, 2)->(2, 3)->(2，5)-> (1, 5)-> (1, 6)-> (1, 5)->(1, 6)-> (1, 7)-> (2, 7)-> (2, 8)-> (2, 9)-> (3, 9)->(4, 9)

(4, 0)->(3, 0)->(2, 0)->(1, 0)->(1, 1)->(0, 1)->(0, 2)->(0, 3)->(0, 4)->(0, 5)->(0, 6)->(0, 7)->(0, 8)->(0, 9)->(1, 9)->(2, 9)->(3, 9)->(4, 9)

(4, 0)->(3, 0)->(3, 1)->(3, 2)->(3, 3)->(3, 4)->(3, 5)->(3, 6)->(3, 7)->(3, 8)->(3, 9)->(4, 9)

Q-learning:

(4, 0)->(3, 0)->(3, 1)->(3, 2)->(3, 3)->(3, 4)->(3, 5)->(3, 6)->(3, 7)->(3, 8)->(3, 9)->(4, 9)

当ε为 0.0000001 时，两种方法输出结果稳定，均为:

(4, 0)->(3, 0)->(3, 1)->(3, 2)->(3, 3)->(3, 4)->(3, 5)->(3, 6)->(3, 7)->(3, 8)->(3, 9)->(4, 9)

除此之外，当ε较大时，Sarsa 的一个 epoch 也往往很长，导致代码运行时间较长。