第四次大作业画图说明

第一张图画q-learning学习曲线，反应出累积奖励的变化。

x轴是 episode number, y轴是累积奖励。

注1：由于训练所需的episode number总数较大，纵坐标可以画每隔一段episodes的平均累计奖励，例如每1000个episodes的平均累积奖励（具体间隔需根据自己的实验参数自行调节）。

注2：可以尝试不同训练参数，尝试动态变化的Q值更新步长和探索系数。

第二张图画最终学出来的最优值函数 v\*(s)。

x轴和y轴分别对应状态空间的两个维度，z轴是各个状态对应的最优值函数v\*(s)。

注3：该图为3d图，python中3d图作图方法可参考plot\_wireframe

<https://matplotlib.org/3.1.1/gallery/mplot3d/wire3d.html>

另外的图不硬性要求，同学们也可以增加一些额外的图来辅助说明实验过程和实验结果。