宏变量（超参数等）

任务自己设计table结构，填写测试数据完成#2中两个模块的编写

#1.状态-动作table状态（车速度位置、杆速度位置）简化到2个（暂时忽略速度）

#2.模块epsilon贪心 A=max(a) Q(S,a) 这两种不太相同 前者更像考虑单步，后者考虑了后续的（不过后续的也是估计出来的，符合采样思想 而是动态规划那种已知模型的）

A的选择

if

随机选action

else

选最优action

Q(S,A)的（增量）更新（记录进入Q(S,A)的次数）

公式

n可看作调节Qn (S, A) 与reward+…Q(S’, A’)间关系的量，n可以理解成以前尝试过的次数，n越大、新增的reward+…Q(S’, A’)看起来就越渺小，1/(n+1)可看作表征新增的reward+…Q(S’, A’)的重要程度的

Qn+1(S, A) = (nQn(S, A) +reward+…Q(S’, A’))/(n+1)

= Qn(S, A) +1/(n+1) (reward+…- Qn(S, A))

for episode

for step

end for

end for

#for 循环能否变为矩阵运算加速 numpy对于循环有没有加速的方案

车位置 杆位置 动作

由状态个数（编号）、动作个数（编号）直接生成Q-table

Q-table类 getter/setter方法直接封装成函数

状态是一个向量、状态容量可能未知、但Q-table一定要可扩展

import numpy as np

arrType=np.ones(1, n)\*2

arr=np.zeros(arrType)

print(arr.shape)