实验报告

2017202130 杨春陶

Openai gym 的cartpole是一个控制小棒使其不落下的一个游戏，求怎样移动使200步内award最高。

Observasion为返回的一个四维向量，代表小车在轨道上的位置、杆子与竖直方向的夹角、小车速度以及角度变化率。

在这里我采用的是模拟退火算法来模拟出最佳策略，通过对这个向量求加权和，于是便可以根据加权和值决定采取相应的动作。

首先生成一个随机四维向量和一个误差，进行测试得分。

之后根据award值调整各权值，因为没有固定的调整方向，随意设置了按正态分布产生的差向量调整权值。

这样迭代10000，根据爬山算法的条件停止算法。

按照爬山算法的测试结果并不理想，因为是无方向的调整权值，导致和随机数再一次生成权值相比，效果虽提升了，但平均也需要迭代1700多次才能找到最优。也是因为无方向的调整和实验次数多了以后翻山的可能性逐渐降低，导致常常陷入局部解。在多维向量中无方向的爬山最为致命。