## 人工智能作业总报告

摘要：通过对本学期的学期，对人工智能的知识有了初步的了解。同时按照课程要求，根据上课所学完成了相应的实验。本文将对本学期所作三次实验进行回顾，并对所有完成的实验进行总结。

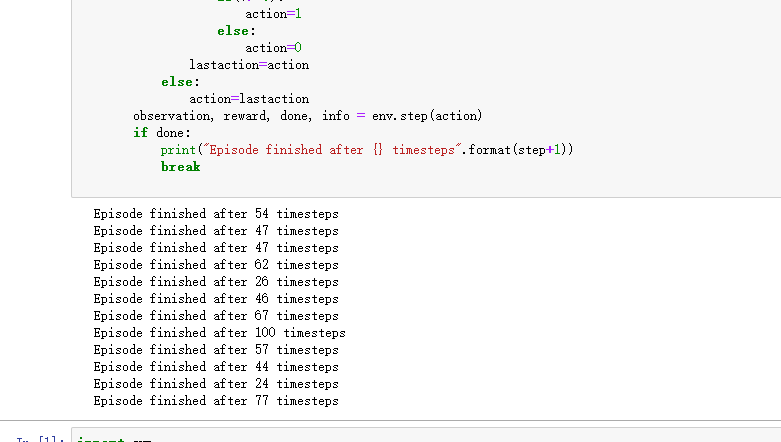
关键词：CartPole-v0，机械学习，深度学习，规则

在本学的学期之中，课程从人工智能的发展史开始，到机械学习，深度学习。详细地介绍了关于人工智能的知识。并设置了相应的实验，本次报告争对的实验是通过不同方法完成gym库中的CartPole-v0游戏，使得游戏得分更高。CartPole-v0游戏又称平衡杆或者倒立摆游戏，游戏是通过控制一辆小车的左右移动使得车上的竖杆不倒，并且小车不触碰两边的边界。游戏中返回了四个参数小车的速度，杆的角度，小车的位置，以及杆的速度。根据四个返回的参数给出小车下一步前进方法使得得分最高便是实验的关键。

第一次实验

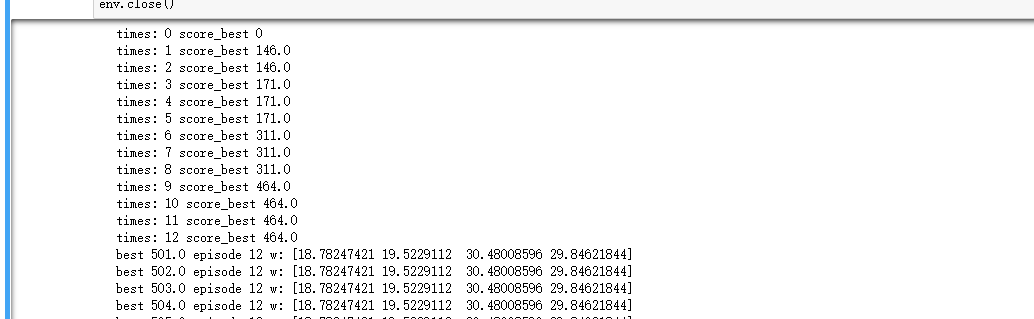
实验要求：基于规则完成游戏

实验思路：第一次实验要求采用基于规则的方式进行游戏，由于要设置规则就需要利用到每次返回的参数值，根据参数值进行判断下一步前进的方向。一个简单的尝试：通过对角度进行判断，当角度在一定范围之内，即木棒接近垂直的时候左右两边随意晃动，当木棒角度超出范围的时候，对角度，角度变化率以及车的速度进行一次方向上的判断，如果角度和角度变化率不同号的时候则保持原来的速度方向，当两者同向的时候，则对速度的方向进行更改。之后的速度方向就和第一次判断的相同，当角度变小到一定范围时候就重复之前的步骤。简单来说就是设置角度变化的阈值，与之内随意变化，当角度超出范围之后，就寻找使得角度变小的方向。这个规则简单也易于实现，实验结果如下图所示：



看到实验的结果波动很大，平均在50次左右说明该规则并不好。通过对返回的参数数据的观察发现，分数较低的情况都会出现角度变化率在一段时间后会增加很大，导致当对小车方向进行修正时已经来不及了，因此失败。

规则不行，又去尝试了基于贪心思想的模拟退火算法对游戏进行了多次尝试得到结果：



结果明显优于按照规则进行实验的实验结果。

第二次实验：

实验要求：基于机械学习完成游戏

实验思路：该次实验要求使用机械学习的方法来让我们的游戏进行更加的智能。由于游戏需要我们决定的是小车的移动方向，并且方向是按照返回的参数值决定的。因此，可以得到结论：每一次的参数值决定我们下一步的步骤。这个显而易见，但是很重要，因为我们可以根据返回的参数值进行分类来得到下一步的选择。当然分类的依据是返回的参数值，而划分的类别包含小车的两个移动方向0和1，及左和右。当完成分类器训练之后，接下来的游戏便可根据小车当前的参数值使用分类器进行训练，便可以得到下一步的选择了。

因为需要对分类器进行训练因此需要适量的数据，而这数据必然是好的数据才值得使用。因此利用上次实验中模拟退火的方法来进行游戏来获取数据，因为模拟退火过程中有机会得到很高的分数，是之前实验之中唯一可以产生高分结果的算法。当然数据是有要求的，这里使用两种数据：得分大于500和得分大于1000分。各有1000条左右的数，因为每一次生成的数据只是一次结果的（得分满足情况），其他情况被舍弃，生成结果存在txt中，之后需要自己对数据处理合并保存到csv中等待之后使用。

数据格式为：下一步行动，返回参数（4个数值）

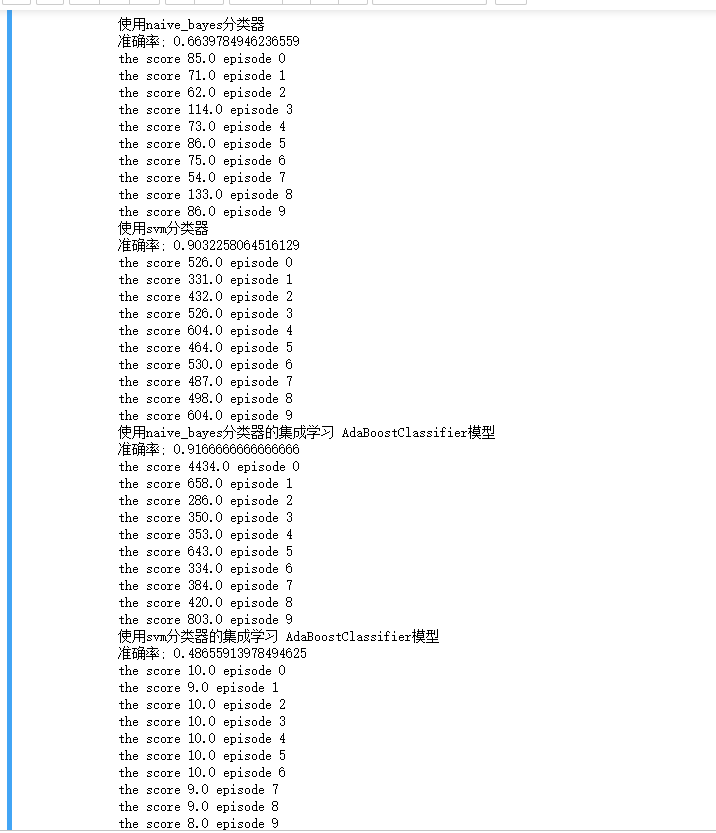
0 -0.020597992176613057 0.01031230735352364 -0.030507967044375407 0.03267692909412155

根据思路和准备得到实验步骤为：使用数据生成代码生成数据，手动处理数据得到分类器可以训练的数据，采用不同的分类器进行分类（该次实验主要采用贝叶斯和svm分类器和集成学习模型AdaBoostClassifier，AdaBoostRegressor），使用分类器进行游戏得到结果。

在实验中实验中的参数，不管是对训练集的划分还是对AdaBoostRegressor参数的设置都是进行过手动调试，保证每个分类器的准确率都在最大值附近，排除了参数对分类器影响的这个无关变量。

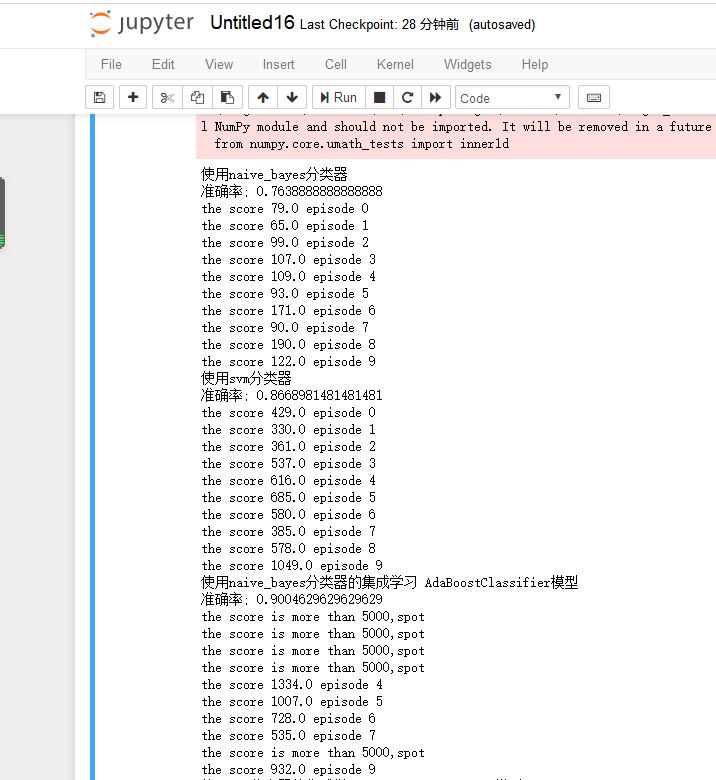
实验结果：

得分大于500分的数据集实验结果：





得分大于1000分的数据集实验结果：





结果说明与分析

1. 对比两种分类器的差距，争对该次实验可以清楚的看到相同的数据贝叶斯分类器分类准确率明显小于svm分类器分类的准确率，这对之后的游戏影响很大，导致肉眼可以的得分差距。因此可以看到贝叶斯分类器效果差于svm。
2. 对比在集成学习模型AdaBoostClassifier上使用svm和贝叶斯又得到截然不同的答案，贝叶斯的集成学习模型使得分类的准确率明显提高的许多，在90%附近。而svm分类器在集成学习模型下准确率不仅没有提高反而严重下滑。因此两者得分又出现巨大的差异。
3. 对于AdaBoostRegressor使用了500个弱的分类器，可是准确率却一般和贝叶斯分类器效果一样但是进行游戏时候得分又十分的差，还比不上贝叶斯分类器的结果。
4. 对于不同数据集对最后结果的影响。不同数据集对于准确率较高svm和贝叶斯集成学习模型影响较大，对于其他三种情况影响几乎可以忽略。而这些影响在于得分的上下界限，当用得分更好的训练集，上下界限都得到不错的提升，即分数的平均情况明显提升，并且得分大的时候可能会带来意外的分数，在贝叶斯集成学习模型中得分有许多次超过5000分的情况（大于5000分的情况被限制不出现了，因为程序可能长时间不停，难以等到结果）。这些都说明了得分更高的数据集效果更好，但只针对哪些准确率较高的，对于较低准确率的情况数据集影响较小。

第三次实验：

实验目的：基于深度学习完成游戏

实验思路：该次实验要求使用的深度学习的方法使得游戏更加智能。在实验二中曾使用分类器来实现游戏，在深度学习中，同样使用神经网络搭建了一个分类器来尝试解决问题。相对于机械学习的数据要求，搭建神经网络训练时使用的数据更加麻烦一点，因此对之前收集的数据重新进行了编排。此外除了使用神经网络的分类器还使用了策略梯度和q-learning的算法去实现该游戏。

对于神经网络搭建的分类器需要数据的准备，而对于策略梯度和深度学习的算法则需要更多的时间来训练模型得到一个较好的模型。训练数据格式；

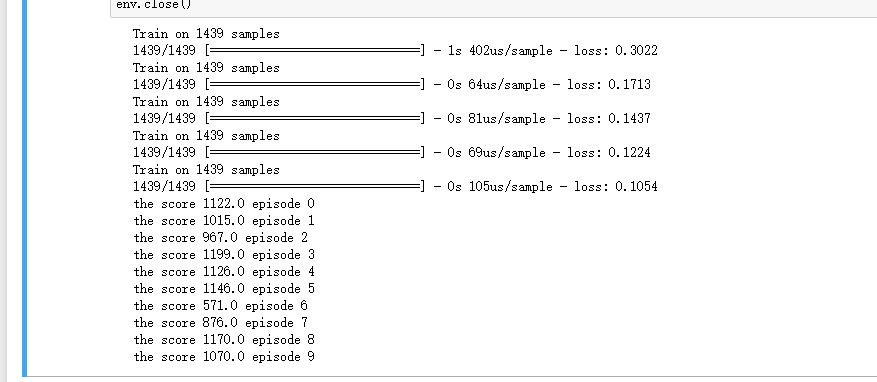
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| -0.020597992 | 0.010312307 | -0.030507967 | 0.032676929 | 1 | 0 |

前四位是每次返回的参数，而之后的1，0作为一个二维数组来决定移动方向，因为模型预测出的结果是也是一个二维数组表示两个方向出现的概率，因此[1,0]表示向左，[0,1]表示向右。对预测结果选择概率大的作为方向的选择。

三个方法实现说明和结果：

1.分类器实现：

搭建一个 4 x 64 x 20 x 2 的网络，输入层为4，输出层为2。对数据集进行训练得到数据集，得到模型之后对行为进行预测得到结果。

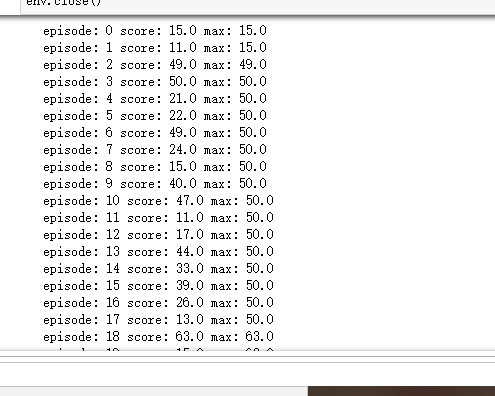


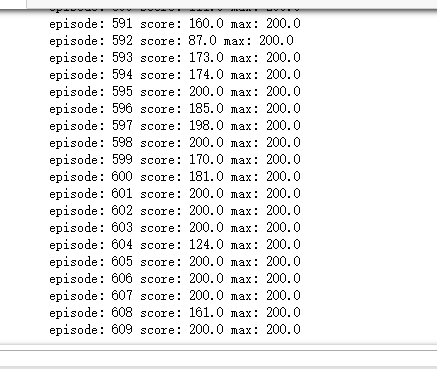
我对模型进行了五次训练可以看到，每增加一次训练loss会变得更小，模型的效果就会好一些，下面的数据是对模型对不同观测行为的预测结果的得分情况，可以看到结果分数平均在1000分左右可以看到模型效果不错。

1. 策略梯度算法

Policy Gradient 网络的输入也是状态(State)，输出是每个动作的概率。例如 [0.7, 0.3] ，这意味着有70%的几率会选择动作0，30%的几率选择动作1。所以导致了算法有更多的不确定性。该次实验搭建的神经网络位输入层为4，输出层为2，隐藏层为100。

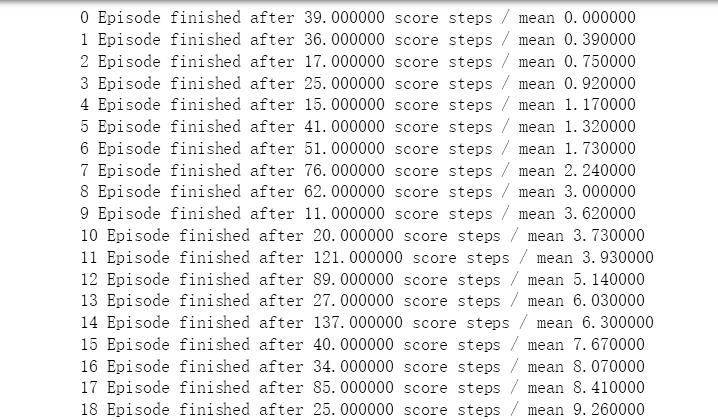
优化策略：累加期望的衰减，某一步的累加期望等于下一步的累加期望乘衰减系数gamma（0.95），加上reward。这使得得分越高之前步骤期望值就越大，说明该步骤出现的概率就越大，因此我借助损失函数来向期望大的方向优化。对于累加期望大的动作，可以放大loss的值，而对于累加期望小的动作，那么就减小loss的值。这样便得到优化策略，之后对游戏进行反复训练，在得分高于一定情况时便结束游戏。

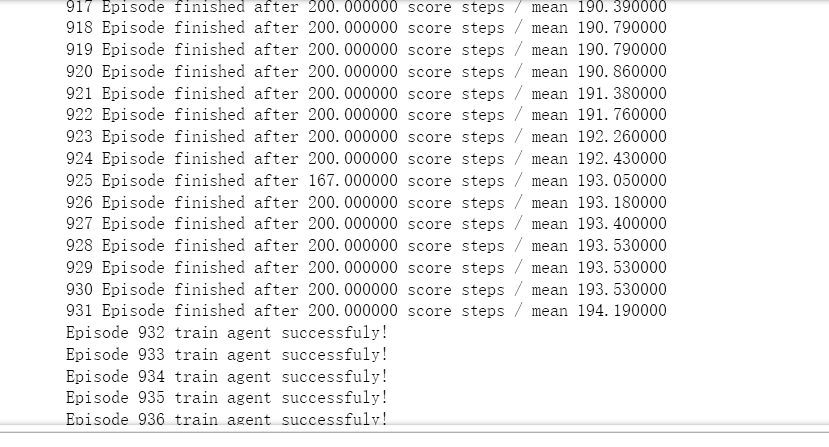


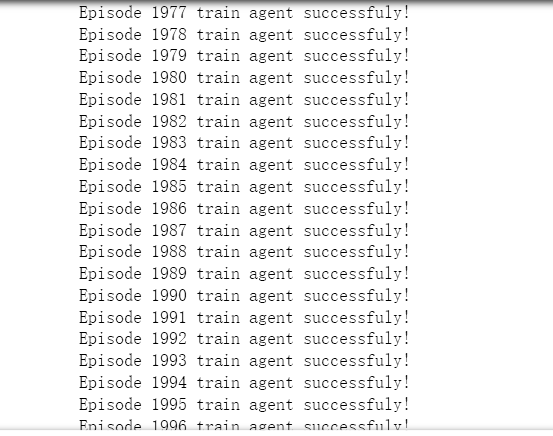


可以看到训练到600多次便可接近200分，分数也稳定在100以上，在经过更长的训练结果会更好。

1. q-learning 算法
2. Learning 就是要学习在一个给定的 state 时，采取了一个特定的行动后，能得到的奖励是什么。在这次实验中给定的状态就是小车的4个状态值，行动就是向左向右，奖励就是得分。之后对每一个步骤进行学习，来更新q-table。下一步行动就从表格中选择。但这样还是难以取得好的结果，因此引入了惩罚机制和贪心策略让算法更加智能。具体在代码中实现。







实验共进行了2000次游戏，当近100场游戏平均得分大于195时成功，可以看到开始时候每次游戏分很低，随着时间进行分数逐渐升高至稳定，由于学习模型最高分限制在200，如果放开限制那么模型在预测时随着时间进行效果应该会更好得分更高。

实验总结：

在第一次实验之中通过规则来完成游戏得到的结果都带有不确定性，即使采用贪心算法来实现游戏得到的结果也不是稳定的，随着游戏次数的增加一些超出规则的情况和预期之外的错误都会发生，导致游戏结果较差，因此仅靠规则是难以实现智能的。第二次实验通过机械学习来实现游戏，机械得到的结果远远好于基于规则实现的结果，然而机械学习需要获得好的结果就必须在学习前给定优质的数据和选择好的机械学习的方法，这些都对最后的结果至关重要，然而数据的获取同样限制了机械学习，没有好的数据来进行训练也就难以实现智能。第三次实验采用了神经网络与强化学习的方法来进行游戏，对于神经网络的分类器仍然需要数据的准备，而对于策略梯度算法好的策略进行学习，强化学习则需要结合其他算法的思想或者结合深度神经网络来达到更好的结果。对于强化学习和策略梯度算法不需要数据的准备却需要大量的学习时间，时间越长程序也会变得更加智能。这解决了机械学习中需要优质数据的缺陷。实验帮助自己更好的学习和吸收课程中涉及到的人工智能知识，同时不同的算法让自己更加充分地了解人们对人工智能地探索过程和不同时期人们对智能思考。随着人们不断探索，算法不断改进，人工智能终将完全实现。