# 第二次实验报告

**实验目的:**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。

**实验准备**：因为需要对分类器进行训练因此我们需要数据，而这数据必然是好的数据才值得我们使用。因此我利用上次实验中使用模拟退火的方法来进行游戏的代码来获取数据，因为模拟退火过程中有机会得到很高的分数，是我之前算法之中唯一可以产生高分结果的算法。当然数据是有要求的，这里我使用两种数据：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，因此没准备太多数据。但实验结果还行。

数据格式为：下一步行动，返回参数（4个数值）

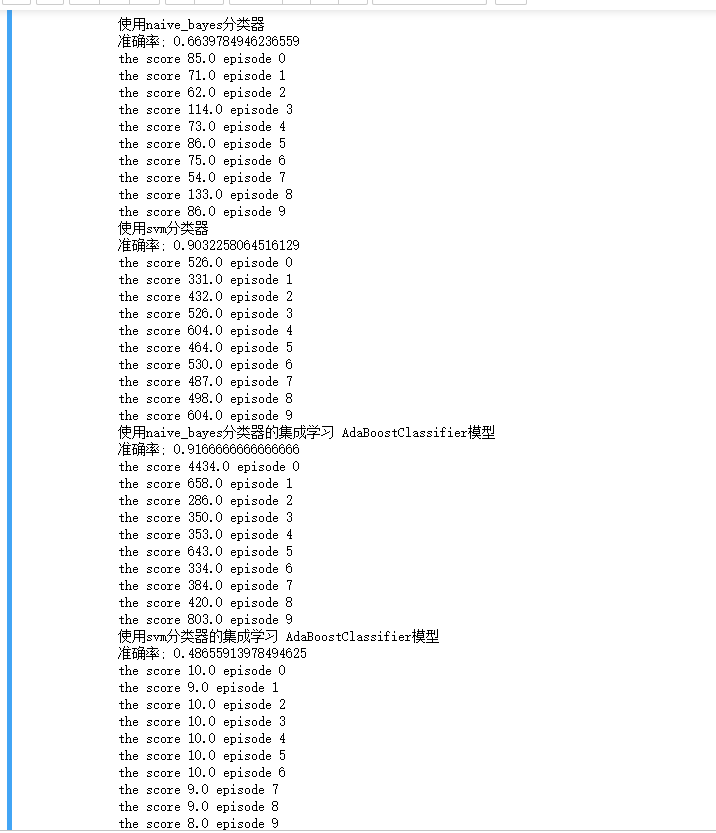
0 -0.020597992176613057 0.01031230735352364 -0.030507967044375407 0.03267692909412155

**实验上交内容**：数据集得分大于500的数据集（500.csv），数据集得分大于1000的数据集（1000.csv），两个python代码，一个是产生数据集代码（2017202116第二次实验2.py），一个是进行游戏的代码（2017202116第一次实验2.py），一份实验报告。

**实验步骤：**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2中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。

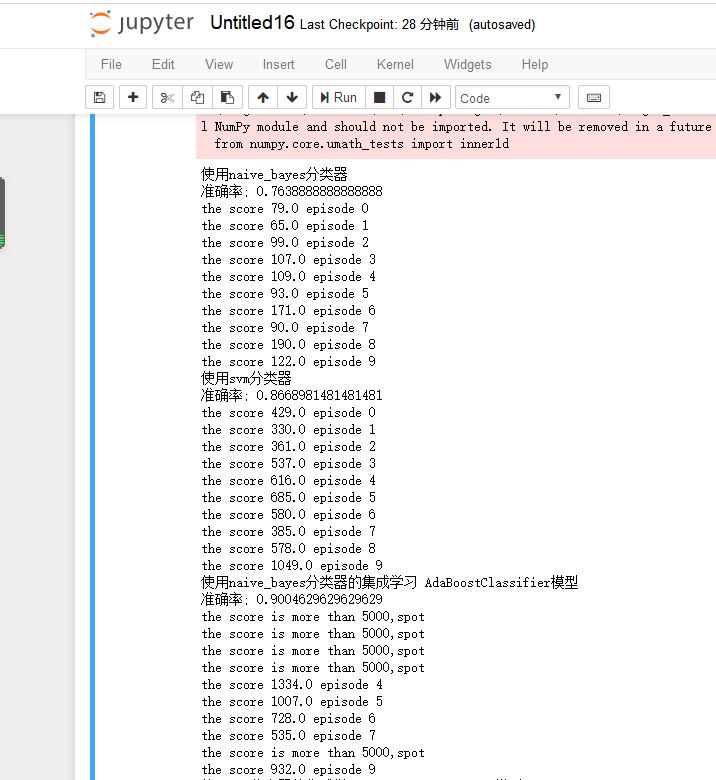
**实验截图（每个分类器进行十次游戏）：**

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：





**结果分析：**

说明：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。

1. 对比两种分类器的差距，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（不知道为什么）。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，可是准确率却一般般和一般的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（这里有点困惑摸不清楚为什么）。
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分的情况被我限制不出现了，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。

**实验中的问题：**

该次实验遇到的问题不多，只要是对实验结果的一些疑问，比如在之前结果分析中提到的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和对于AdaBoostRegressor准确率和一般的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。这两点有疑问，其余问题都不大，在实验中顺利解决。

**实验小结**

该次实验加深了自己对机械学习的理解，尤其是在于分类器方面，感觉自己对分类器有了一个更深层次的理解。同时学会了怎么使用不同的分类器，怎么进一步提升分类器的效果。收获还是不错的。但是实验还有许多可以改进的地方，首先是数据的获取和数据量的大小，这一点在该次实验中是需要进行改进的，一是数据集可以更好，二是数据量应该更大这样效果将会更好，其次是分类器的优化，该次实验只是使用了两种不同的分类器，还有几种可供我们选择的分类器模型可以使用，因此还有许多值得优化的地方。