实验八 决策树

1. 实验目的
2. 熟悉Python的基础使用，会调用.txt文件

2.理解和掌握决策树的基本原理和实现过程

3.掌握信息熵、信息增益的计算

二、实验内容

1.了解决策树原理

2.运行决策树程序

3.移植程序，用决策树解决实际问题

三、实验原理：见课本。

**任务1：手动编程,输出决策树模型**

1. 打开原文链接：<https://www.jianshu.com/p/de0ad1c793d6>
2. 运行文中程序，理解决策树工作原理，会计算信息增益和信息熵；
3. 根据输出结果，画出决策树；
4. 改变数据集，解决经典的打网球问题，画出决策树。

注意，样本数少时可以手动输入，当样本数较多时，应使用批量读取数据，可参考任务2中的.txt文件读取。

{'outlook': {'overcast': 'yes', 'sunny': {'humidity': {'high': 'no', 'normal': 'yes'}}, 'rainy': {'windy': {'FALSE': 'yes', 'TRUE': 'no'}}}}

问题描述：今天是否去打网球（play）主要由天气（outlook）、温度（temperature）、湿度（humidity）、是否有风（windy）来确定。样本中共14条数据。

NO. outlook temperature humidity windy play  
1 sunny hot high FALSE no  
2 sunny hot high TRUE no  
3 overcast hot high FALSE yes  
4 rainy mild high FALSE yes  
5 rainy cool normal FALSE yes  
6 rainy cool normal TRUE no  
7 overcast cool normal TRUE yes  
8 sunny mild high FALSE no  
9 sunny cool normal FALSE yes  
10 rainy mild normal FALSE yes  
11 sunny mild normal TRUE yes  
12 overcast mild high TRUE yes  
13 overcast hot normal FALSE yes  
14 rainy mild high TRUE no

**任务2：直接调用决策树库，进行分类**

参考原文链接,完成以下任务。

https://blog.csdn.net/zhuiyunzhugang/article/details/105544014

1. 手动生成打网球的.txt文件(可直接复制)；
2. 运行后如果报错显示没有pydotplus模块，可以不安装，就把程序中“4.把决策树结构写入文件”下面对应的代码屏蔽掉，不影响结果；
3. 逐条解释运行结果；
4. 解释程序中def函数的作用；

（5）详细解释下列三行语句的作用(包括各个参数)；

①data=np.loadtxt("play.tennies.txt",delimiter="",dtype=str, converters={0:outlook\_type,1:temperature,2:humidity, 3:windy,4:play\_type} )

②x, y = np.split(data,(4,),axis=1)

③x\_train, x\_test, y\_train, y\_test = train\_test\_split(x, y, test\_size=0.3)

(6)添加以下两行程序，运行，写出输出结果，并解释。

from sklearn import metrics

print(metrics.classification\_report(y\_test, answer))

**任务3：实际应用**

结合西瓜数据集2.0，任选一种方法，生成决策树模型或分类结果，和课本结果做比较。

