**《人工智能导论》**

**实验四 指导书**

北京工商大学人工智能学院

# 实验四 分类：决策树与SVM

## 一、实验目的

学习并掌握信息熵的计算过程；

掌握决策树的ID3算法原理以及优缺点；

熟悉并运用scikit-learn工具包的模型。

## 二、实验任务

利用Python编程，完成实验内容中要求的题目，并完成实验报告，提交至Bb平台。

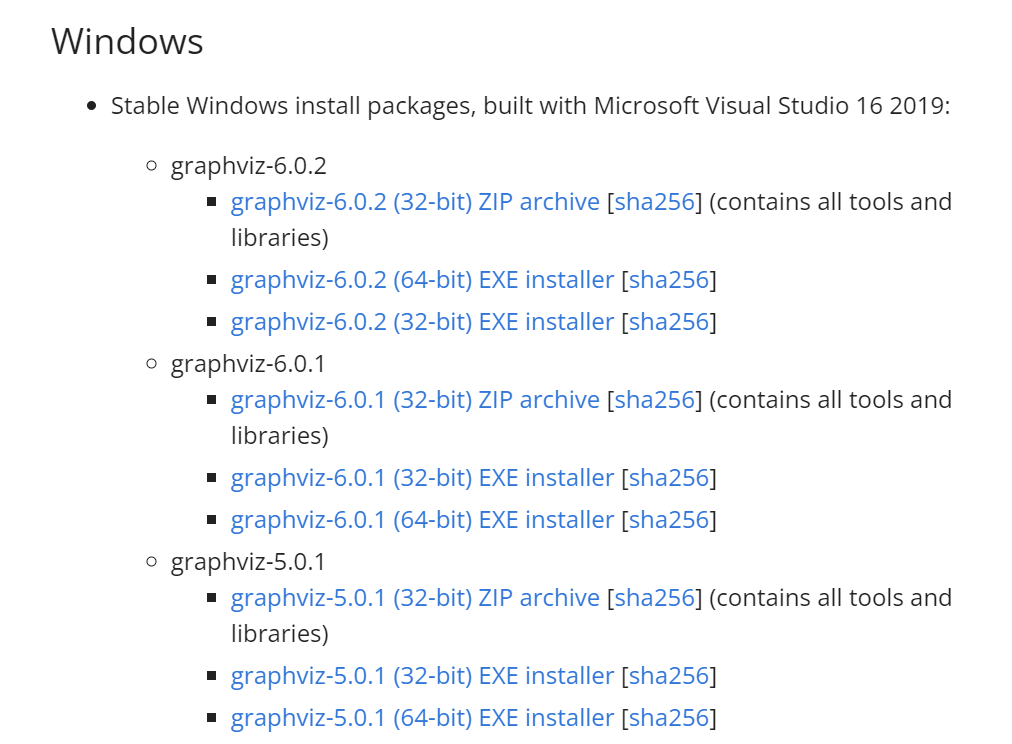
## 三、实验内容

### 1、对于不同的人是否会购买电脑，已在DecisionTreeSet.csv文件中给出。请根据文件中提供的数据，假设决策树是一个二叉树，用ID3算法生成一个决策树，并通过图的形式，或字典嵌套的方式，例如{‘outlook’:{‘rain’:{‘wind’:{‘weak’: ‘yes’,‘strong’: ‘no’}},‘sunny’:{‘normal’: ‘yes’,high: ‘no’}}}，输出决策树的结构。

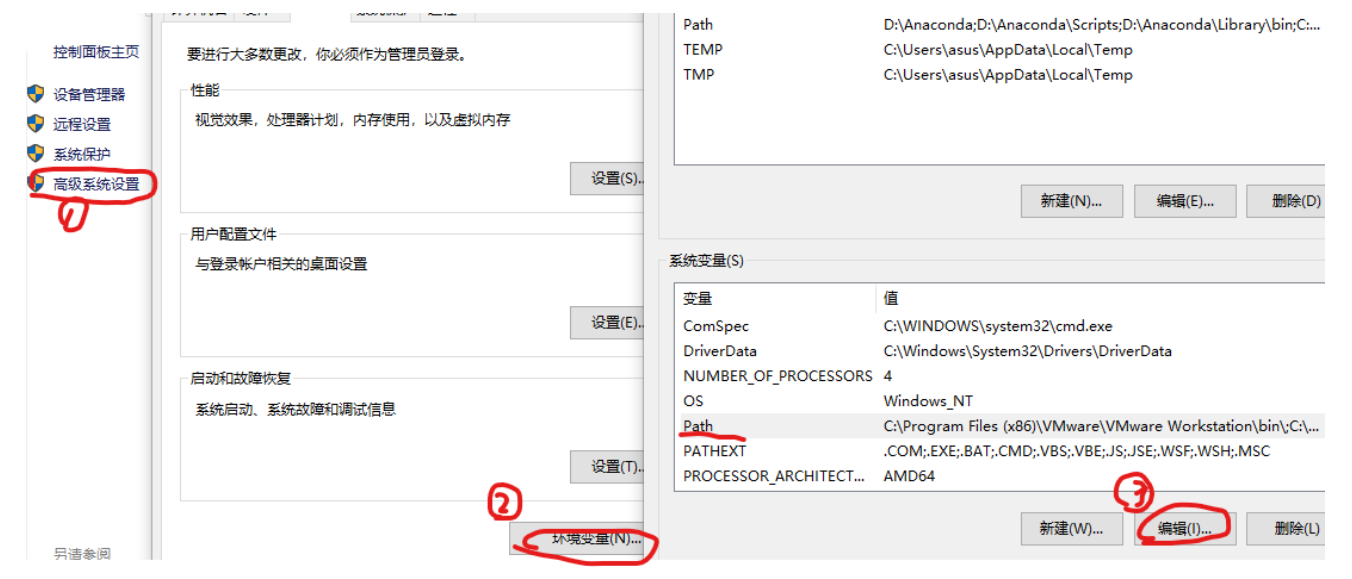
提示：输出图形式的决策树可以安装Graphviz，安装步骤如下：

（1）打开命令提示符输入：pip install graphviz

（2）到网站<https://graphviz.org/download/> 下载安装包EXE，注意选择操作系统。



（3）安装完成后添加环境变量。右键电脑属性 高级系统设置 环境变量 编辑系统变量 选择新建，把 C:\Program Files\Graphviz\bin 加进去（替换为自己的路径）



图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

（4）完成之后，可以运行如下代码，如果不行，可以重启pycharm

import graphviz

dot\_data = tree.export\_graphviz(model,

out\_file = None,

feature\_names = [替换为特征名称]

class\_names = [替换为类别],

filled = True,

rounded = True,

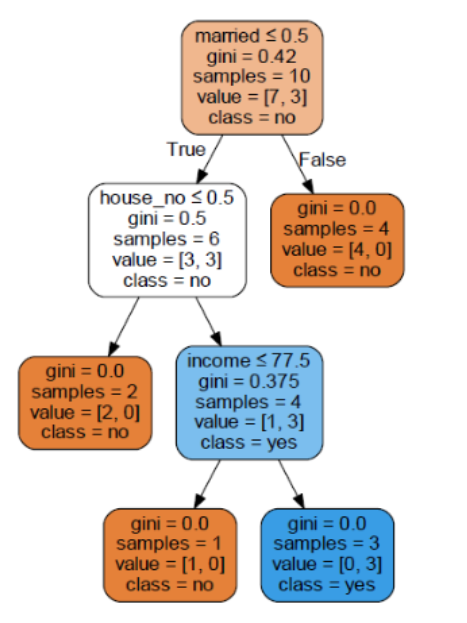
special\_character = True)

graph = graphviz.Source(dot\_data)

graph.render('filename')

这里的model为创建的scikit-learn库中的决策树模型。

最后生成如下图所示的决策树。



也可以使用其他方法输出图结构的决策树，不局限于这一种。

### 2、iris数据集是常用的分类数据集，请用SVM对该数据集进行两两分类，比较不同的核函数对分类结果准确率的影响，并做出数据的散点图和超平面。

提示：

iris数据集一共有150个样本（数据集的行数），有4个属性（数据集的列数）：Sepal Length（花萼长度），Sepal Width（花萼宽度），Petal Length（花瓣长度），Petal Width（花瓣宽度）。种类：Iris Setosa（山鸢尾）、Iris Versicolour（杂色鸢尾），以及Iris Virginica（维吉尼亚鸢尾）。

sklearn库可以直接导入数据集，代码如下：

from sklearn import datasets

iris = datasets.load\_iris()

## 四、实验要求

1. 认真编写并上机调试本实验的程序。

2. 实验报告要求：

1. 实验报告内容包括但不限于：实验目标、实验原理、实验内容、实验设计流程图、实验结果截图、实验总结等；
2. 根据实际情况写出本实验的实验目的；
3. 根据实验具体情况指出本实验的实验内容；
4. 按要求完成实验任务；
5. 设计实验流程图；
6. 具有实验结果（程序运行的截屏图或代码）
7. 实验总结（收获、体会等）。
8. 实验报告提到电子版到Bb平台。