

## 实验二 机器学习实验

## 实验任务2-1 决策树实验

对scikit-learn工具包中的鸢尾花（iris）数据构建决策树实现分类  
要求：

- 1、实现教材上给出的决策树构建算法（ID3算法），对iris数据进行分类
- 2、注意鸢尾花数据的特征为连续特征，教材上的算法只能处理离散特征，因此，在使用之前需要将鸢尾花的连续特征转换为离散特征，例如可以转换为“长、中、短”等。
- 3、将你构建的决策树分类器的分类结果与sklearn工具包中提供的决策树（参考提供的源代码DecisionTree.ipynb）进行比较，并进行结果分析。
- 4、实验完成后提交源代码，并将实验内容写入实验报告。

## 示例：使用Sklearn库自带的鸢尾花数据集生成一棵决策树

见源代码 `DecisionTree.ipynb`

在anaconda里使用graphviz，用于显示决策树

- (1) 下载网址：[www.graphviz.org](http://www.graphviz.org)
- (2) 选择最新的可执行文件进行graphviz安装
- (3) 设置环境变量Path为 “安装目录\bin”
- (4) 用管理员权限运行anaconda
- (5) 在命令窗口执行 “conda install python-graphviz”  
进行安装

## **sklearn.tree.DecisionTreeClassifier**

- 1、Scikit-Learn使用的DecisionTreeClassifier采用的是CART算法
- 2、因为DecisionTreeClassifier采用的CART训练算法是随机的。因此，即使在相同的数据上训练相同的模型，每次也可能产生不同的决策树。
- 3、DecisionTreeClassifier具体的使用方法参考：  
<https://scikit-learn.org.cn/view/784.html>

## 实验任务2-2 K-均值聚类实验

### 参考源代码 Clustering.ipynb

对使用scikit-learn工具包中make\_blobs函数生成的模拟数据进行聚类分析  
要求：

- 1、实现教材上给出K均值聚类算法，对模拟数据进行聚类。
- 2、将你的聚类结果与使用sklearn工具包中提供的KMeans聚类结果进行比较分析。（源代码Clustering.ipynb提供了运用sklearn的KMeans进行聚类的方法，并提供了可视化代码，你可以在你自己的代码中使用这些可视化代码）
- 3、实验完成后提交源代码，并将实验内容写入实验报告。

## 提交文档（三个文件）

- 1、决策树实验源代码文件，文件名举例：[谢羿衡-2023302061163-实验二-DTree.ipynb](#)
- 2、K均值聚类实验源代码文件，文件名举例：[谢羿衡-2023302061163-实验二-KMeans.ipynb](#)
- 3、实验报告（模板与实验一相同），文件名举例：[谢羿衡-2023302061163-实验二.pdf](#)