# 实验一 分类技术---鸢尾花分类

## 实验内容

采用所学分类算法，对鸢尾花数据进行分类，具体要求如下：

1.填补鸢尾花数据集中的缺失数据，并划分训练集、测试集

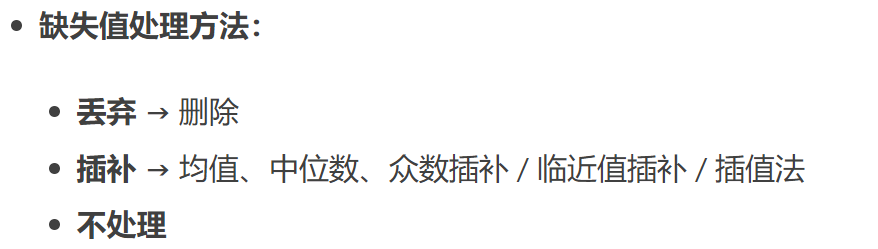
2.采用决策树、KNN、朴素贝叶斯对数据分类，比较其在测试集上的预测准确率；

3.可视化决策树模型，打印出特征属性的重要性。

注：实验一和二的数据在https://www.jianguoyun.com/p/DVOLxIIQqrn6CRip5JUE

## 分析及设计

要求1.



选用中位数插补方法

划分训练集、测试集

用load\_iris（），train\_test\_split（）

要求2.

DecisionTreeClassifier（）

KNeighborsClassifier（）

GaussianNB（）

## 详细方案

决策树分类及可视化



KNN分类



朴素贝叶斯

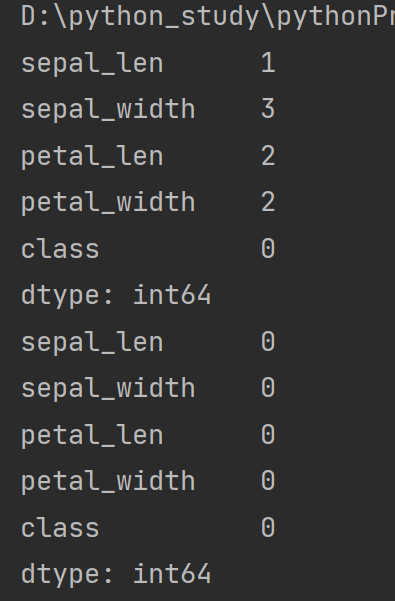


代码形式基本相同，故不重复注释

## 4.实验结果

填补数据，输出缺少值的个数，确保填补成功



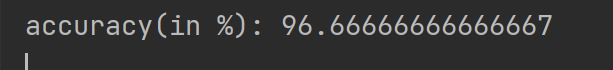


准确率

决策树：



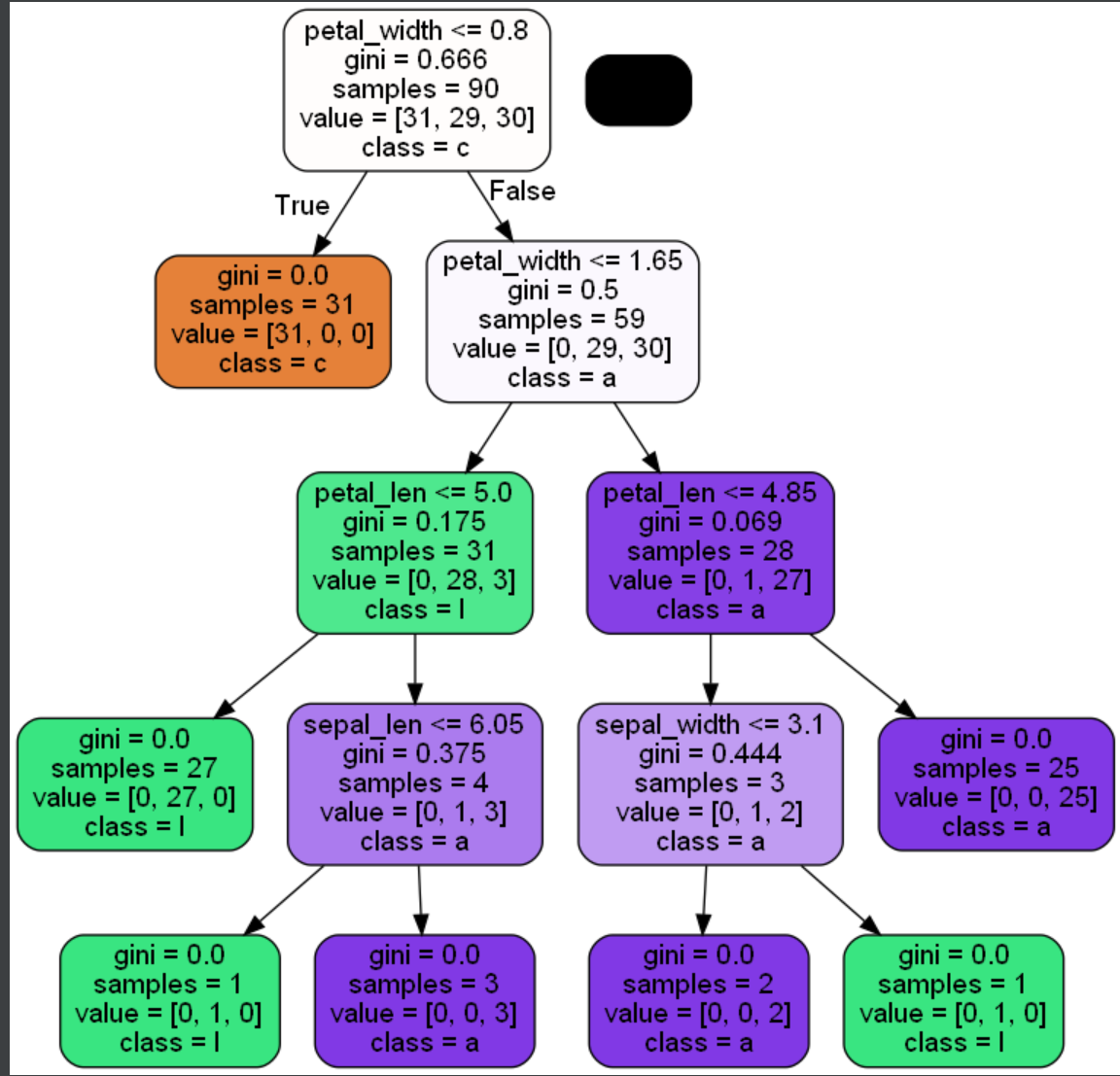
KNN：



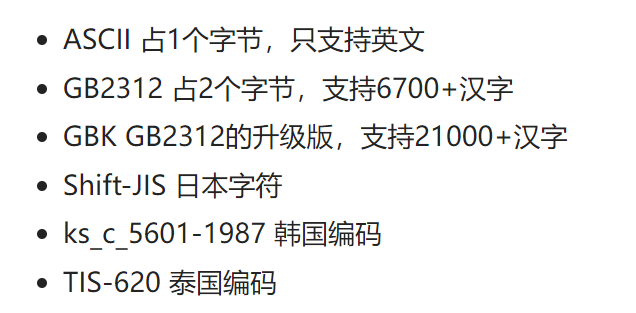
朴素贝叶斯：



决策树



编码问题：



需使用GBK

## 5.心得体会

加深了对理论知识的理解，增强了动手实践能力。