|  |
| --- |
| 西南民族大学学生实验报告  教学单位：计科学院 实验室名称：BS226 实验时间：2019年10月6日 |
| 实验项目名称：实验2—决策树 实验成绩： 教师签名： |
| 实验项目报告内容（1、实验目的；2、实验主要过程与结果；3、分析讨论）。  **一、实验目的**  学会运用决策树处理分类问题。  **二、实验主要过程与结果**   1. 决策树的构造    1. 信息增益   在trees.py中增加如下代码，如图1所示，在主函数中加入语句：  myDat, labels = createDataSet()  print(myDat)  print(calcShannonEnt(myDat))  运行结果如图2所示。    图 1    图 2   * 1. 划分数据集   编写一个函数，使算法能够按照给定特征划分数据集，程序清单如图3所示，在主函数中加入如下语句：  myDat, labels = createDataSet()  print(myDat)  print(splitDataSet(myDat, 0, 1))  print(splitDataSet(myDat, 0, 0))  运行结果如图4所示。  编写一个能选择最好的数据集划分方式的函数，如图5所示，在主函数中加入如下测试语句：  myDat, labels = createDataSet()  print(chooseBestFeatureToSplit(myDat))  print(myDat)  运行结果如图6所示。    图 3    图 4    图 5    图 6   * 1. 递归构造决策树   定义一个能返回出现次数最多的分类名称的函数，如图7所示，定义一个能创建树的函数，如图8所示，在主函数中添加如下代码：  myDat, labels = createDataSet()  myTree = createTree(myDat, labels)  print(myTree)  运行结果如图9所示。    图 7    图 8    图 9   1. 在Python中使用Matplotlib注解绘制树形结构    1. Matplotlib注解   新建treePlotter.py文件，使用文本注解绘制树节点，代码如图10所示，在控制台中输入：  treePlotter.createPlot()  treePlotter.createPlot()  结果如图11所示。    图 10    图 11   * 1. 构造注解树   在treePlotter.py中添加获取叶子节点数目的函数、树的层次的函数和预存树的信息的函数，如图12、图13、图14所示。  在控制台的输入和输出结果如图15所示。  更新代码    图 12    图 13    图 14    图 15  更新代码，如图16、图17、图18所示。  在控制台输入图19所示代码，绘制出的树形图如图20所示。    图 16    图 17    图 18    图 19    图 20   1. 测试和存储分类器    1. 测试算法：使用决策树执行分类   在tree.py中添加分类函数，如图21所示。  在主函数中写入如图22所示语句：  运行结果如图22所示。    图 21    图 22    图 23   * 1. 使用算法：决策树的存储   使用pickle模块存储决策树，函数如图24所示。  在主函数中写入如下语句：  myTree = treePlotter.retrieveTree(0)  storeTree(myTree, 'classifierStorage.txt')  grabTree('classifierStorage.txt')  运行后同级目录下生成classifierStorage.txt文件    图 24   1. 示例：使用决策树预测隐形眼镜类型   在主函数中读取lenses.txt，使用决策树预测隐形眼镜的类型，代码如图25所示。  运行结果如图26所示。    图 25    图 26  **三、分析讨论**   1. 在使用Matplotlib注解时，出现了错误：Backend Qt5Agg is interactive backend. Turning interactive mode on.   错误原因：matplotlib的默认backend是TkAgg，而FltkAgg, GTK, GTKAgg, GTKCairo, TkAgg , Wx or WxAgg这几个backend都要求有GUI图形界面的，所以在ssh操作的时候会报错。  解决办法：指定不需要GUI的backend(Agg, Cairo, PS, PDF or SVG)，在plt.show()之前，加上plt.switch\_backend(‘agg’)。   1. 在使用pickle模块存储决策树时，出现了错误：write() argument must be str, not bytes   错误原因：fw这个变量的类型应该是bytes的  解决办法：将'w'改为'wb'。  编译后提示UnicodeDecodeError: 'gbk' codec can't decode byte 0x80 in position 0: illegal multibyte sequence错误，是由于之前保存的txt文件是用二进制写入的所以在读取时也应该用二进制，即读文件中的fr = open(filename)应该改为fr = open(filename, 'rb')。 |

注：实验报告的内容及格式可由学院根据学科专业特点确定；全校各专业必须使用学校统一封面。