|  |
| --- |
| 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9月28日 |
| 实验项目名称：实验1—KNN算法 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**   1. 了解KNN算法； 2. 了解KNN算法的应用。   **二、实验主要过程与结果**   1. k-近邻算法的概述    1. 准备：使用Python导入数据   在kNN.py中增加如下代码，如图1所示（注：在python3中“from numpy import”应改为“import numpy as np”，且在之后用到numpy库中的函数时，要加上“np.”）。  定义主函数，如图2所示，在文件末尾加上如图3所示语句，运行kNN.py，运行结果如图4所示。    图 1    图 2    图 3    图 4   * 1. 实施kNN分类算法   编写一个函数，使用kNN算法将每组数据划分到某个类中，如图5所示。  在main函数中加入语句：print(classify0([0, 0], group, labels, 3))，运行结果如图6所示。    图 5    图 6   1. 使用k-近邻算法改进约会网站的配对效果    1. 准备数据：从文本文件中解析数据   将文本记录转换为Numpy的解析程序如图7所示，在main函数中加入语句：datingDataMat, datingLabels = file2matrix('datingTestSet2.txt')  print(datingDataMat)  运行结果如图8所示。    图 7    图 8   * 1. 分析数据：使用MatPlotlib创建散点图   在main函数中加入如图9所示语句，运行所得图形如图10所示。    图 9    图 10   * 1. 准备数据：归一化数值   归一化特征值的函数如图11所示，在main函数中加入如下语句：  normMat, ranges, minVals = autoNorm(datingDataMat)  print("normMat:\n", normMat)  print("ranges:\n", ranges)  print("minVals:\n", minVals)  运行结果如图12所示。    图 11    图 12   * 1. 测试算法：作为完整程序验证分类器   分类器针对约会网站的测试代码如图13所示，在main函数中输入如下语句：datingClassTest()  运行的部分结果如图14所示。    图 13    图 14   * 1. 使用算法：构建完整可用系统   约会网站的测试函数如图15所示，在main函数中加入如下语句：  classifyPerson()  运行，输入数据，结果如图16所示    图 15    图 16   1. 手写识别系统    1. 准备数据：将图像转换为测试向量   新建一个handwritingClassify.py文件，导入kNN。将图像转换为测试向量的函数如图17所示，在main函数中输入如下语句：  testVector = img2vector('testDigits/0\_13.txt')  print(testVector[0, 0: 31])  运行handwritingClassify.py，结果如图18所示。    图 17    图 18   * 1. 测试算法：使用k-近邻算法识别手写数字   测试代码如图19所示，在main函数中输入：  handwritingClassTest()  运行结果如图20所示。    图 19    图 20  **三、分析讨论**   1. k-近邻算法是使用距离测量的方法来分类物品的； 2. 不一定k越高，错误率越低，可以通过找到合适的k值； 3. 在准备数据这一步骤，必须把待处理数据的格式转换成分类器可以接受的格式。 |

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。