计算机学院实验报告

|  |  |  |
| --- | --- | --- |
| 实验题目：基于kNN的手写字体识别 | | 学号：201600181073 |
| 日期： | 班级： 智能16 | 姓名： 唐超 |
| **Email：[1984386166@qq.com](mailto:1984386166@qq.com)** | | |
| 实验目的：  基于Python实现kNN算法实现手写字识别，然后使用sklearn库的kNN算法实现手写字识别，掌握sklearn机器学习库的使用。 | | |
| 实验环境介绍：  Python3.6 & sklearn & jupyter notebook | | |
| 解决问题的主要思路：  本实验基于kNN算法，主要算法思路与实验五一致，具体在实验步骤中说明。 | | |
| 实验步骤：  1.准备数据集。trainingDigits为训练数据集的文件，文件数量为1935；testDigits为测试数据集的文件，文件数量为947。每个文件的命名格式都为“真实数字\_编号.txt”。为了简便处理，实验中，用txt文本文件表示图片。原图片中像素值为黑色（0,0,0）的像素点在txt中对应的用0表示，像素值为白色（255,255,255）的像素点用1表示。所以，你只需要处理这些文本文件即可，不用再去解析图片格式。每个txt文件中，数据共有32行和32列，这是由于原图片的大小为32X32。  2.加载手写字训练集数据。由于kNN核心算法中，每个点都是用向量表示的。尽管已经转换为文本形式了，不过图片数据依然是32X32的二维数据格式，首先需要将其转换为一维数组，即表示成一个向量。在kNN.py中添加函数img2vector用于将32X32的二维数组转换为一维数组，该函数传入参数文件名，返回转换后的一维数组。  3.验证一下我们的kNN手写字识别算法的准确性。测试使用的数据集位于testDigits下。统计错误的数量，计算错误率。  **4.** 基于sklearn的kNN实现手写字识别**。**使用sklearn.neighbors.KNeighborsClassifier实现kNN算法。填入KNeighborsClassifier函数相关参数，对测试集数据进行分类。  关键实验代码见指导书。 | | |
| 实验结果展示及分析：  最终结果，测试总数据量为946，不使用用sklearn，设置k=3，出错的数据量为11，错误率为0.0116。使用sklearn最终结果出错的数据数量为13，错误率为0.0137。 | | |