**Lab 6** 实验题目

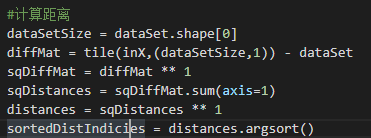
|  |  |  |  |
| --- | --- | --- | --- |
| 学号 | **17363011** | 学院 | 专业 |
| 姓名 | **陈政培** | **智能工程学院** | **智能科学与技术** |

1. 实验目的： **利用python实现KNN分类器**
2. 实验环境：vs code、python3.6.8
3. 实验步骤：
4. Knn1就是最简单的根据距离排序得到的聚类结果

* 我们在knn.classify测试分类器时，调整了用于分类的输入向量，当输入为【0，1】时输出结果为B，输入为【0，1.1】时输出结果为A。证明分类器正常工作

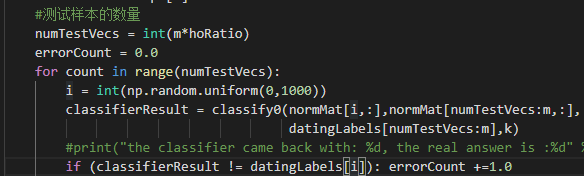
1. Knn2改进约会网站的配对效果

* 根据knn2教程编写了对datingTestSet2.txt的分类，分为’not at all’ ,’in small doses’ ,’in large doses’三个种类并通过knn2.classifyPerson测试结果
* 为了方便我们调试不同k值下对错误率的影响我们对代码进行了一点更改，直接在knn2.py文件中为datingClassTest和classifyPerson函数加入了一个参数k，并且给定了原文件中的预设k值
* 只需要在主函数文件调用对应函数时传递不同的k值即可观看效果
* 让knn分类器使用曼哈顿距离进行计算，需要在knn2.py中的classify0函数中修改计算距离的代码



将平方项和开根项改为1即可

* 为了探索随机选取训练样本，对错误率的影响需要在knn2.py中的datingClassTest函数中修改测试样本的代码



把原来按顺序循环的变量i变成有numpy提供的random函数从1000个样本中获取到的随机数

* 测试不同样本数目对错误率的影响就比较简单只需要修改knn2.py中的datingClassTest函数中测试样本比例参数hoRatio的大小

1. 使用knn识别手写体

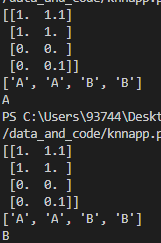
* 实例中只提供了knn3.py函数，并没有给出主函数操作，所以需要在主函数中自主调用
* 因为knn3.py中createDataSet函数和trainingDataSet函数并没有区别，且handwritingTest函数中包含了调用createDataSet训练分类器的语句，所以主函数仅仅需要调用

knn3.handwritingTest()

1. sklearn实现knn2和knn3

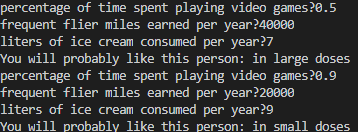
* knn2直接重写了一套程序，主函数和函数都在sklearnknn2.py文件中
* knn3则是仿照knn3.py，把在classify过程中直接调用KNN分类器

1. 实验结果与分析：
2. Knn1实验结果



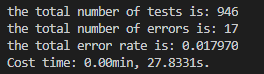
1. Knn2实验结果

* 分类测试



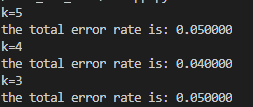
结果符合理论，游戏时间少里程多的人更加有吸引力

1. Knn3手写体测试结果



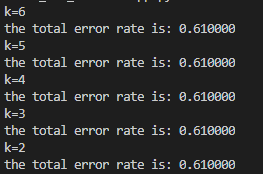
错误率仅有1.797%

1. 作业：
2. k值对错误率的影响



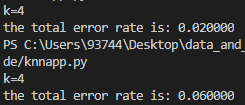
K=4时，分类器错误率达到最低

1. 使用曼哈顿距离对错误率的影响

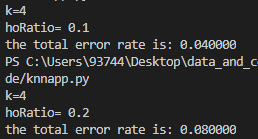


K值将不会再产生影响，使用曼哈顿距离错误率比欧氏距离要高

1. 随机选取训练样本，测试不同样本数目对错误率的影响



同样的代码随机选取样本，两次运行结果错误率不同



取消随机样本后，调整测试样本比例参数，测试样本比例越大，错误率越高

1. 将knn1，knn2，knn3中的语句

from numpy import \*

修改成

import numpy as np

并让代码正常运行

只需要在所有numpy提供的函数前加上np. 引用即可

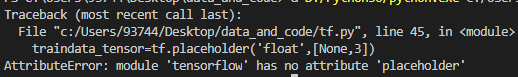
主要是zeros，tile，array等函数

1. 使用sklearn实现knn2和knn3效果上由于机制原理相同没有什么差异，但是使用sklearn不再需要自己编写算法，可以更加简单方便的解决问题
2. 程序代码完成了，但是tensorflow出现了问题无法正常运行
3. 实验总结：

* 本次实验初期使用的是最新版本的python3.8.0，在安装matplotlib、sklearn等库时出现安装失败的现象。回退老版本python3.6或3.7系列即可解决问题
* 上几次实验用到matplotlib绘图的部分无法正常显示汉字问题得到解决，只需要在程序开头加上

plt.rcParams['font.sans-serif']=['SimHei']

* 在用tensorflow实现knn的时候遇到了问题



无法在tensorflow中找到placeholder，这个问题没有得到解决

参考文献：

1. <https://blog.csdn.net/asialee_bird/article/details/81051281>
2. <https://www.cnblogs.com/CXZzero/p/10747636.html>
3. https://www.jianshu.com/p/ab296440b0de