**西安电子科技大学**

**安全前沿讨论班（I） 课程实验报告**

**实验名称 K近邻算法的实现**

学 院 网络与信息安全学院 班 级 1618019

姓 名 曹寅峰 学 号 16020610025

实验日期 2019 年 5 月 21 日

|  |
| --- |
| 指导教师评语：  指导教师：  年 月 日 |
| **实验报告内容基本要求及参考格式**   1. 实验目的   应用已经掌握的Python程序设计语言的相关知识，大致了解如何用Python实现K近邻分类算法，并创建一个自己的KNN（K-Nearest Neighbor）分类器。对所设计的问题进行编程实现，从而达到对所学知识的练习和巩固。   1. 实验环境   Python3.5 jupyter   1. 实验基本原理及步骤（或方案设计及理论计算）  计算两点间距离 根据两点间距离公式，直接返回即可。注意这样写的运算单位是单个坐标，很方便。 找出最高频率项 这里我们创建一个频率字典，键值为类别-票数，同时利用count找出票数的最大值对应的键，最后返回这个键 另一种高效方法 利用ss库可以高效找出频率最大的项，其中第一个[0]为频率最大的项的列表形式，第二个[0]为列表中的这一项的元素形式。 找到临近邻居 这里我们先计算p和points中的每个元素的距离，并进行升序排序，返回前K个最邻近邻居的**索引号** Knn函数 利用前面产生的序列，对应outcomes的分类结果，再从中选出频率最高分类结果，完成预测 产生模拟数据 这段代码的意思是产生0+-1区间的5个2维元素。  随后我们产生了50个二维随机点，并随机分为2类，并将图像存储到pdf中。 学习怎样预测网格; 学习enumerate的使用； 学习怎样使用numpy中的meshgrid. 这里是对空间中每个点（网格）进行knn预测。其中enumerate() 函数用于将一个可遍历的数据对象(如列表、元组或字符串)组合为一个索引序列，同时列出数据和数据下标,画出预测网格。  Meshgrid用于创建指定个数的**坐标矩阵。最后利用这些信息，预测结果保存到prediction\_grid中。** 学习怎样画出预测网格 其中主要代码是  plt.pcolormesh(xx, yy, prediction\_grid, cmap = background\_colormap, alpha = 0.5)  将之前的坐标矩阵，预测结果矩阵，和背景颜色绘出图像。 在真实数据集上应用自己创建的KNN分类器 从iris中选出二维向量，并且画出他们的坐标，分类。  再利用同样的方法，画出预测网络  这里可以看到，我们的预测结果和官方预测结果稍有不同  我们的预测结果正确概率是96%  官方均方误差为83.3%  而我们的均方误差为84.7%，稍大一些   1. 实验结果分析及回答问题（或测试环境及测试结果）   两点间距离计算正确  最高频率项也正确    找出了最邻近的点  预测结果为1  总结：  通过这次实验，了解了knn分类器的原理，以及numpy的编程技巧，mpplot画图的技巧。 |