**实验四 K近邻分析与应用**

【实验目的】

1 理解K近邻分类算法的基本原理。

2 掌握K近邻分类算法的编码实现方法。

3 理解K近邻分类算法的优劣势。

【实验类型】

设计型

【实验学时】

2学时

【实验环境】

Windows 7以上操作系统

Python3.0以上版本

Pycharm开发环境

Spyder开发环境

【实验要求】

K近邻分类方法的基本思路是：不对预先的训练数据集进行任何处理，当测试一个新来的数据属于哪个类时，计算下它与训练集中所有实例的距离，从中选出K个距离最小（近）的实例，统计下这K个实例属于哪个类的最多，就将新来的数据归为哪一类。

谚语：“走路像鸭子，叫起来像鸭子，看起来像鸭子，那它很可能就是鸭子”。

举例来说，如果我们有下面的训练集，其中每个实例有5个属性，最后一列标出了它们属于哪一类？

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 编号 | 属性1 | 属性2 | 属性3 | 属性4 | 属性5 | 所属类 |
| 1 | 1.1 | 1.2 | 1.3 | 1.4 | 1.5 | A |
| 2 | 2.1 | 2.2 | 2.3 | 2.4 | 2.5 | B |
| 3 | 3.1 | 3.2 | 3.3 | 3.4 | 3.5 | A |
| 4 | 4.1 | 4.2 | 4.3 | 4.4 | 4.5 | A |
| 5 | 5.1 | 5.2 | 5.3 | 5.4 | 5.5 | B |

假设现在有一个新数据（2.11，2.21，2.22，2.23，2.24，2.25），我们取K = 3，在训练集中与它距离最近的三个实例肯定是实例1，2，3，其中两个属于A类，一个属于B类，那我们就将新数据归为A类。

请使用python编写程序实现KNN算法并使用提供的数据集调试验证，最后将实验报告（包含附录源代码）上交到QQ作业。

读取数据和数据点间距离计算的方法请参考：

