



**数据挖掘**

**课后作业**

|  |  |
| --- | --- |
| 专 业： |  |
| 班 级： |  |
| 学 号： |  |
| 姓 名： |  |

2024年09月23日

## 课后作业

1. 简述如何计算被如下属性描述的对象的相异性：

（1）标称属性；

（2）非对称的二元属性；

（3）数值属性。

2. 假设有一个描述仓库货品的二维数据集：

|  |  |  |
| --- | --- | --- |
|  | A1 | A2 |
| x1 | 1.5 | 1.7 |
| x2 | 2 | 1.9 |
| x3 | 1.6 | 1.8 |
| x4 | 1.2 | 1.5 |
| x5 | 1.5 | 1.0 |

该数据集包含了2个数值属性，5个数据对象。假设你的任务是根据用户提供的A1和A2上的描述，给出最相似的商品。请回答以下问题：

（1）假设用户提供的描述为（1.4，1.6），分别使用欧氏距离、曼哈顿距离、上确界距离和余弦距离给出查找结果（完整过程）。

（2）假设公司将该数据集修改成大规模数据集，且每个数据对象都是高维数据，请分析欧氏距离、曼哈顿距离、上确界距离和余弦距离在该数据集上的优缺点，给出你认为最合适的相似性度量方式并给出理由。

## 回答区域

**提交入口：**

<https://send2me.cn/h4t5rIpK/SD-IN8l-H02GAQ>