**第十四章读书报告**

09118223 吴亦珂

读书进度：十四章第一节

1. **问题列表**
2. （我提出）聚类算法一般在什么情况下使用？具体有哪些使用实例。

讨论结果：一般用于数据没有标签的情况下进行，通过聚类可以得到数据的特点火堆数据进行一定的处理。聚类的应用实例也很广泛，比如说数据挖掘以及模式识别。

（别人提出）聚类和降维代表了从两种方向进行压缩的方法，那么这两种方法之间有什么相似性和差异？

讨论结果：聚类是指将不同的实体分成一个一个类别，所以是对不同实体进行操作，而矩阵储存数据时，一般每一行一个特征，每一列一个样本，因此相当于对实体个数进行操作，所以书上称为对纵向结构进行操作。降维是指对于数据的特征数进行操作，降低数据的特征数，因此相当于对横向结构进行操作。

（别人提出）什么是无标注数据？他和有标注有数据的差异体现在哪里？

讨论结果：无标注数据就是原始的数据，即数据没有标签。所以对于这些数据在使用的过程中还需先对其进行分类再加以研究。

（别人提出）P257页的14.6式中的协方差矩阵按定义好像是n\*n的，但是这样矩阵乘法的长宽就对不上了？

讨论结果：这个主要是用书上14.13式子，根据该式，协方差矩阵应该是mm的。这样与矩阵乘法规则相符。

1. **下周读书计划**

下周计划读完第十四章，聚类在之前的课程也有所学到，同时也进行过代码的实现，所以我准备这周结合书本再次复习一遍聚类算法。

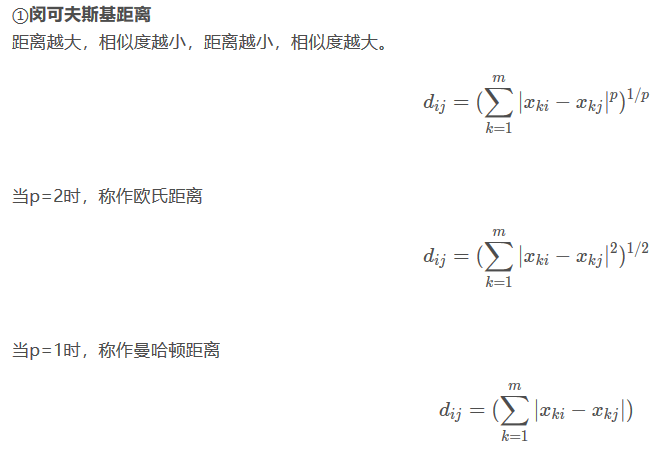
1. **读书收获**

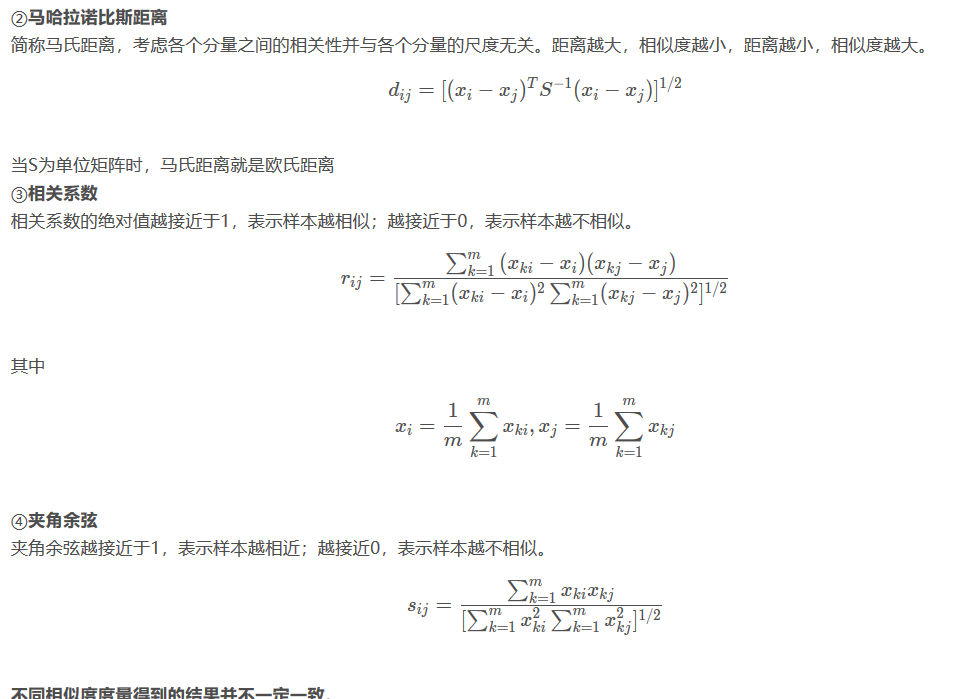
第十二章主要对监督学习方法进行了总结，介绍了各种学习算法的使用问题，模型以及学习的策略。

第十三章主要对无监督学习进行了简要地概述。首先介绍了无监督学习的基本原理，从数据的横向结构以及纵向结构角度说明了无监督学习的基本想法。接着介绍了聚类的基本问题：聚类、降维、概率模型估计。最后介绍了机器学习的三要素以及无监督学习的学习方法。

第十四章主要介绍了聚类这一种算法，14.1节主要介绍了聚类的基本概念。

1. 四种衡量相似度的四种方法：

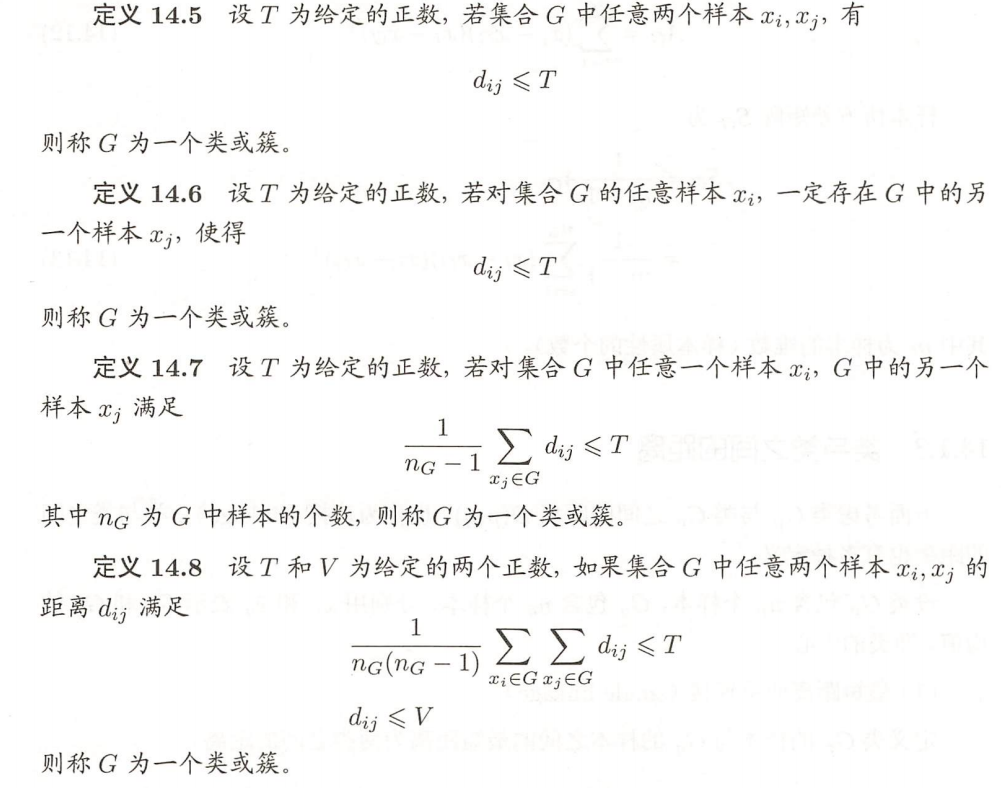




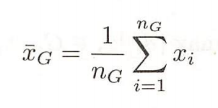
1. 硬聚类：每个样本只属于一个类。

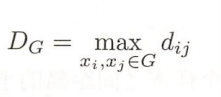
软聚类：每个样本可以属于多个类。

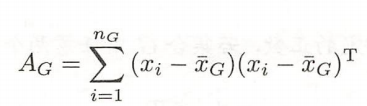
1. 类或簇的四种定义：

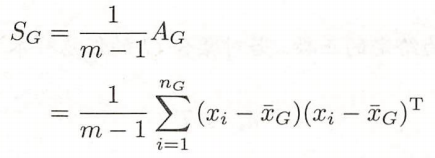


1. 常用的特征：

类的中心，即类的均值：

类的直径，即类中任意两个样本的最大距离：

类的样本散步矩阵：

样本协方差矩阵：

1. 类与类之间的距离：

