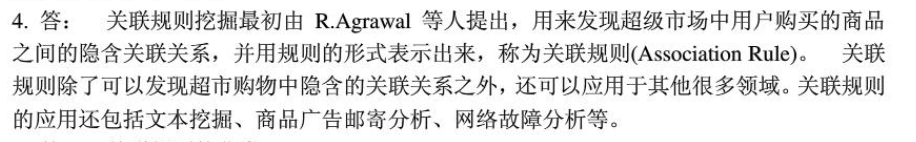
**1.数据仓库就是一个面向主题的、集成的、相对稳定的、反映历史变化的数据集合。(概念，特点)**

**2.什么是关联规则？关联规则的应用有哪些？**

关联规则分析也称为购物篮分析，最早是为了发现超市销售数据库中不同的商品之间的隐含关联关系。成功被应用于市场营销、银行业、零售业、等各个方面。



**3.什么是分类？分类的应用领域有哪些？**

分类是指把数据样本映射到一个事先定义的类中的学习过程，即给定一组输入的属性向量及其对应的类用基于归纳的学习算法得出分类。应用于银行业（辅助工作人员将正常信用卡用户和欺诈信用卡用户进行分类，减小银行损失）、医疗诊断（将正常细胞和癌细胞进行分类）、因特网筛选（协助网络工作人员将正常的邮件和垃圾邮件进行分类）等诸多领域

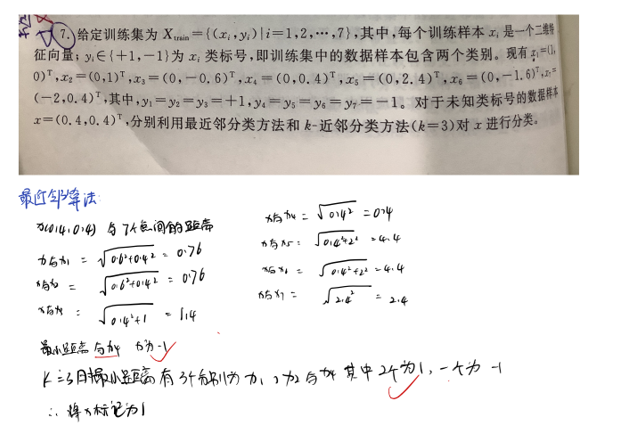
**4.什么是聚类分析？聚类分析的应用领域有哪些**？

将未知类标号的数据集划分为多个类别之后观察每个类别中数据样本的特点并且对某些特定的类别作进一步分析。应用于科学数据分析、商业、生物学、文本挖掘、web数据挖掘等领域

5.分类过程，常见的评价指标

6、关联规则算法。

7. K-means 、kNN基本算法



**8最后一题：300字小作文 与数据挖掘流程有关，列举分类、聚类、关联规则的方法，分析他们的优缺点**

1. **数据挖掘的流程**

1、定义问题

在开始知识发现之前最先的也是最重要的要求就是了解数据和业务问题。必须要对目标有一个清晰明确的定义，即决定到底想干什么。比如，想提高电子信箱的利用率时，想做的可能是“提高用户使用率”，也可能是“提高一次用户使用的价值”，要解决这两个问题而建立的模型几乎是完全不同的，必须做出决定。

2、建立数据挖掘库

建立数据挖掘库包括以下几个步骤：数据收集，数据描述，选择，数据质量评估和数据清理，合并与整合，构建元数据，加载数据挖掘库，维护数据挖掘库。

3、分析数据

分析的目的是找到对预测输出影响最大的数据字段，和决定是否需要定义导出字段。如果数据集包含成百上千的字段，那么浏览分析这些数据将是一件非常耗时和累人的事情，这时需要选择一个具有好的界面和功能强大的工具软件来协助你完成这些事情。

4、准备数据

这是建立模型之前的最后一步数据准备工作。可以把此步骤分为四个部分：选择变量，选择记录，创建新变量，转换变量。

5、建立模型

建立模型是一个反复的过程。需要仔细考察不同的模型以判断哪个模型对面对的商业问题最有用。先用一部分数据建立模型，然后再用剩下的数据来测试和验证这个得到的模型。有时还有第三个数据集，称为验证集，因为测试集可能受模型的特性的影响，这时需要一个独立的数据集来验证模型的准确性。训练和测试数据挖掘模型需要把数据至少分成两个部分，一个用于模型训练，另一个用于模型测试。

6、评价模型

模型建立好之后，必须评价得到的结果、解释模型的价值。从测试集中得到的准确率只对用于建立模型的数据有意义。在实际应用中，需要进一步了解错误的类型和由此带来的相关费用的多少。经验证明，有效的模型并不一定是正确的模型。造成这一点的直接原因就是模型建立中隐含的各种假定，因此，直接在现实世界中测试模型很重要。先在小范围内应用，取得测试数据，觉得满意之后再向大范围推广。

7、实施

模型建立并经验证之后，可以有两种主要的使用方法。第一种是提供给分析人员做参考；另一种是把此模型应用到不同的数据集上