**《数据挖掘》复习题**

**（初稿，2017年）**

一、名词解释

1. 数据挖掘
2. 孤立点分析
3. 数据清理
4. 数据集成
5. 数据归约
6. 数据离散化
7. 数据规范化
8. 关联规则
9. K-均值聚类
10. 文本自动分类
11. 因子分析

二、简答题

1. 互联网数据有哪些主要特征？
2. 简述什么是数据挖掘
3. 简述数据挖掘的主要步骤
4. 什么是孤立点分析？
5. 简述什么是数据清理
6. 简述什么是数据集成
7. 简述什么是数据归约
8. 简述什么是关联规则挖掘
9. 简述什么是聚类分析
10. 简述什么是文本自动分类
11. 举例说明什么是K-均值聚类算法。
12. 简述什么是层次聚类法。
13. 聚类分析有哪些主要方法？
14. 什么是KNN-分类算法？
15. 举例说明什么是决策树
16. 简述数据挖掘的主要软件工具
17. 聚类分析与自动分类有何区别与联系？
18. 因子分析法的主要思想是什么？

三、论述题

1. 详述数据挖掘的主要步骤？
2. 在进行数据挖掘或分析之前为什么要进行数据预处理？有哪些主要方法？
3. 举例说明数据预处理的主要方法？
4. 举例说明什么是关联规则挖掘？它有哪些典型应用？
5. 举例说明什么是聚类分析？它有哪些典型应用？
6. 举例说明什么是自动文本分类？它有哪些主要应用？
7. 什么是数据标准化 ？举例说明有哪些主要方法？
8. 什么是层次聚类法？请举例说明层次聚类法的基本过程。
9. 自动文本分类的一般过程？有哪些主要方法？