

2013 教育部-IBM 产学合作专业综合改革项目 厦门大学《数据挖掘原理及实践》课程习题

第 1 章 引言

1. 讨论下列每项活动是否是数据挖掘任务
 - a) 根据性别划分公司的顾客。
 - b) 根据可盈利性划分公司的顾客。
 - c) 计算公司的总销售额。
 - d) 按学生的标识号对学生数据库排名。
 - e) 预测掷骰子的结果。
 - f) 使用历史记录预测某公司未来的股票价格。
 - g) 检测病人心率的异常变化。
 - h) 监视地震活动的地震波。
 - i) 提取声波的频率。
2. 假定你是一个数据挖掘顾问，受雇于一家因特网搜索引擎公司。举例说明如何使用诸如聚类、分类、关联规则挖掘和异常检测等技术，让数据挖掘为公司提供帮助。
3. 数据仓库和数据库有什么不同？它们有哪些相似之处？
4. 对于如下每个数据集，解释数据的私有性是否是重要问题。
 - (1) 从 1900 年到 1950 年收集的人口普查数据。
 - (2) 访问你的 Web 站点的用户的 IP 地址和访问时间。
 - (3) 从地球轨道卫星得到的图像。

(4) 电话号码簿上的姓名和地址。

(5) 从网上收集的姓名和电子邮件地址。

5. 请简述数据挖掘的基本概念和流程。

6. 简要描述数据挖掘有哪些基本类型的算法以及它们应对的不同数据挖掘任务。

7. 请简述数据挖掘，机器学习和模式识别之间的关系。

8. 举例说明数据挖掘可以面向什么类型的应用。

9. 定义下列数据挖掘功能：特征化，关联，分类，回归，聚类。

10. 数据挖掘系统大致可分为几类？请举出各类的例子。