

## 附录 C 数据处理和挖掘

附录 C: 数据处理  
和挖掘



我们正步入一个数据或许比软件更重要的新时代。

*We're entering a new world in which data may be more important than software.*

——蒂姆·奥莱利 (Tim O'Reilly)

O'Reilly 媒体公司的创始人兼 CEO、Open Source 和 Web 2.0 概念的提出者

### 学习目标

- (1) 了解数据挖掘的基本概念。
- (2) 掌握数据挖掘的基本方法。
- (3) 运用第三方库实现数据挖掘的聚类、分类和回归方法。
- (4) 了解机器学习的基本概念。
- (5) 掌握一种聚类或分类方法。

数据挖掘，又译为资料探勘、数据采矿。数据挖掘就是在大量数据中寻找有意义、有价值信息的过程。当别人空谈故事时，我们要学会用数据说话。在现代社会中，与其求助于个人的信息知识储备，不如借助网络的海量信息筛选来完成特定的任务，这个特定的任务对象并不一定是枯燥的表格，它可以是一颗行星的轨道，可以是云雨的变化，或者是，一朵花？

本章将讲述如何使用 K 均值方法根据花瓣形状将花朵儿们分门别类。

本章以电子资源形式提供，请扫描侧边栏二维码获取内容。

