



数据挖掘

Data Mining

主讲: 张仲楠 教授



廈門大學
XIAMEN UNIVERSITY

课程介绍

为什么需要数据挖掘：

- 数据分析**历史悠久**，我们一直使用数据帮助我们
- 每一次**信息的爆炸增长**会带来巨大影响
- 过往的数据 --- 已经学习过的
- 如今的数据 --- 正在学习的
- 未来的数据 --- 未知的/需要预测的



课程介绍

课程范畴：

- 这是一门什么样的课程？——介绍**数据分析方法**
- 在这门课我们需要解决什么样的问题？
 - 比如：分类与回归、关联分析、聚类分析等
- 我们要学习的工具——python, Jupyter Lab

课程介绍

课程内容：

- 数据挖掘的**基本理论**
- 数据挖掘的**实现过程**
- 常用的数据挖掘**算法**
- 基于python的数据挖掘**应用**

课程介绍

教学进度安排:

各章教学大纲	教学形式	时间安排
<u>第一章 绪论</u> 数据挖掘概念，数据挖掘要解决的问题，数据挖掘的起源，数据挖掘任务等。	课堂讲授，随堂练习，实验	4+2
<u>第二章 数据基础</u> 数据类型，数据质量，数据预处理，相似性和相异性的度量，数据可视化等。	课堂讲授，随堂练习，实验	6+4
<u>第三章 分类</u> 基本概念，决策树分类器，最近邻分类器，朴素贝叶斯分类器，支持向量机，组合方法，模型的过拟合，模型评估，超参数的使用，类不均衡问题等。	课堂讲授，随堂练习，作业，实验	10+6

课程介绍

教学进度安排：

各章教学内容纲要	教学形式	时间安排
<u>第四章 回归分析</u> 基本概念，线性回归，非线性回归，逻辑回归等。	课堂讲授	2
<u>第五章 关联分析</u> 基本概念，频繁项集的产生，规则的产生，关联模式的评估等。	课堂讲授，随堂练习， 作业，实验	3+2
<u>第六章 聚类分析</u> 概述，K均值，凝聚层次聚类，DBSCAN，簇评估等。	课堂讲授，随堂练习， 作业，实验	5+2
<u>第七章 异常检测</u> 概述，异常检测问题的特性，统计方法，基于邻近度的方法，基于聚类的方法等。	课堂讲授	2

课程介绍

先导知识:

- 有一定的概率统计基础
- 有一定编程基础
- 有一定的自学能力

课程介绍

教材:

■ 数据挖掘导论 (第2版)

■ [美] 陈封能 (Pang-Ning Tan) 等著, 段磊 张天庆等 译

■ 出版社: 机械工业出版社

■ ISBN: 9787111631620



课程介绍

课程沟通:

■ Email: zhongnan_zhang (at) xmu.edu.cn

■ Office: 西部片区5#-202

■ QQ 群: 704917850

■ 课程网站:

<https://Int.xmu.edu.cn/course/44710/content#/>



课程介绍

考核办法:

- 平时成绩 10%
- 作业实验 40%
- 期末考核 50%

