



数据挖掘

与

产业实践



▷ - 欢迎 -



朱雪宁

Xuening Zhu

办公室: 子彬院209

邮箱: xueningzhu@fudan.edu.cn



Welcome





- 学习目标 -



- 探索 -

从数据到价值

- 了解 -

基于行业的数据实践

- 学习 -

经典数据挖掘模型

- 磨练 -

团队合作精神



- 学习内容 -

第1讲：数据思维与数据价值	第2讲：数据预处理、描述分析
第3讲：线性回归+逻辑回归	第4讲：逻辑回归+泊松回归
第5讲：简单分类模型	第6讲：树模型（决策树、回归树、随机森林）
第7讲：支持向量机（SVM）	第8讲：聚类分析
第9讲：降维分析（主成分分析+因子分子）	第10讲：文本分析
第11讲：图像分析简介	第12讲：关联分析
第13讲：社交网络分析	

- 学习形式 -

- 整个学习分小组进行，每组5人；同时含有个人作业
- 要求：熟练运用R（或其他）语言 + 数理逻辑分析 + 汇报展示 + 沟通表达
- 作业形式：（1）实际数据；（2）粗糙的背景介绍；
- 作业目标：（1）完成报告；（2）需要汇报的组，PPT演讲稿。
- 平时成绩：
 - 个人成绩：个人作业成绩
 - 小组成绩：随机抽选汇报一次（全班同学打分），每次汇报所有组交一次报告（共4次报告）
- 期末
 - 小组汇报：所有组汇报一次，并提交报告
 - 打一场kaggle比赛，按照排名计算成绩（截止日期：6.30）

- 成绩构成 -

- 每周作业：60%
- 期末报告：40%

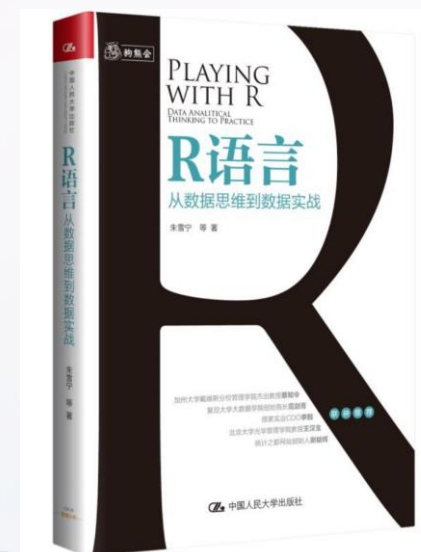
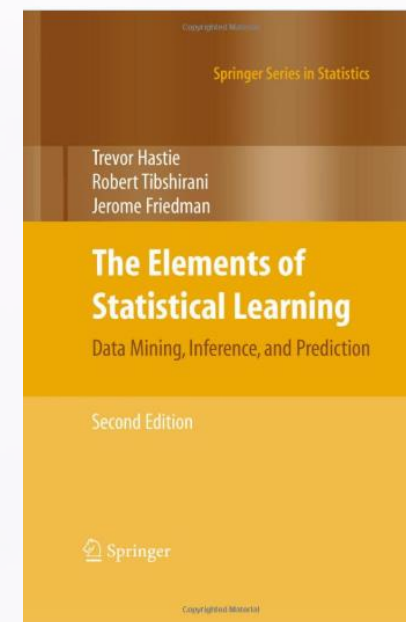
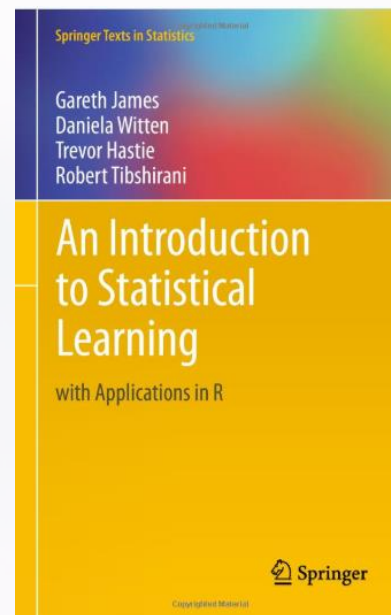
参考
资料

狗熊会



- 参考教材 -

- Introduction to Statistical Learning
- The Elements of Statistics Learning
- 《数据思维实践：从零经验到数据英才》
- 《R语言：从数据思维到数据实战》



Before you learn model building, you should know first.....

“

Every model is **wrong**, but some are **useful**!

– George Box

”



Thank You