MA333 Introduction to Big Data Analysis
Course Introduction

Zhen Zhang

Southern University of Science and Technology

Outlines

Course Info

- Semester 2024-2025 Spring
- Instructor : ZHANG, Zhen （张振）
- Office : Room M5014, School of Science
- Phone : 88018753
- Email : zhangz@sustech.edu.cn
- Office hours : Tuesday morning, 10 :00-12 :00 ; or send email to make an appointment for other time.
- Lecture : 3 credits, 3 hours per week.
- Prerequisite : Calculus (or Mathematical Analysis) ; Linear Algebra ; Probability Theory or Probability Theory and Mathematical Statistics (or other similar courses).

Grading Policy

- Homework : Approximately 6 homework assignments (including programming assignments and written problems). The written homework could be handed in after class.
- In-class quizzes : typically once every two weeks, test how well you learned about the basic concepts, including fill-in-the-blank, single and multiple choices, and simple Q & A
- Programming projects : include coding, data analytics, and reports
- One closed-book final exam
- Grading policy : assignments (30%), quizzes (15%), programming projects (20%), and the final exam (35%).

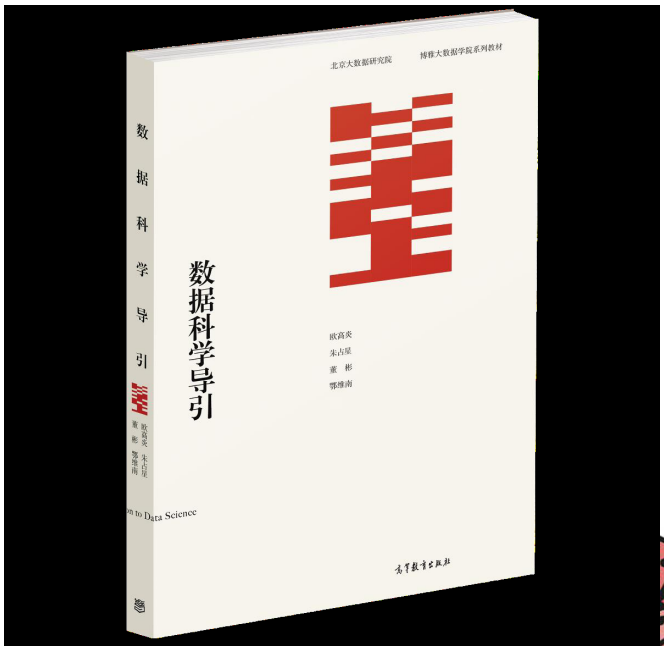
Contents

- Intended for undergraduate students who are interested in pursuing industrial work and research in big data science.
- Concise and self-contained introduction to mathematical aspect of big data science, including **theoretical analysis, algorithms and programming with python**
- Major topics :
  - Introduction to python programming and data preprocessing
  - Three fundamental problems : classification, regression, clustering
  - Model selection, dimensionality reduction
  - Hot topics : text analysis, social network analysis, neural network and deep learning, and recommender systems if time permits

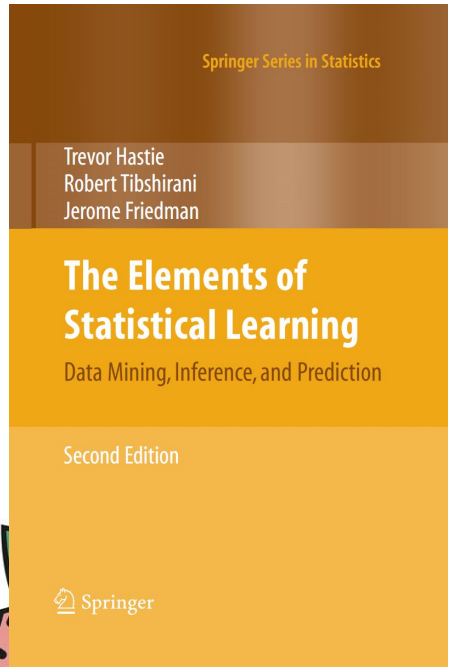
References

- 数据科学导引，欧高炎等著，高等教育出版社，2017.
- 机器学习，周志华著，清华大学出版社，2016.
- An Introduction to Statistical Learning with Applications in Python, by Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani and Jonathan Taylor, Springer, 2023. https ://www.statlearning.com/
- Pattern Recognition and Machine Learning, by Christopher M. Bishop, Springer, 2006.
- The Elements of Statistical Machine Learning : Data mining, Inference and Prediction, 2nd Edition, by Trevor Hastie, Robert Tibshirani, and Jerome Friedman, Springer, 2009. https ://hastie.su.domains/ElemStatLearn/
- Foundations of Machine Learning, by Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar, MIT Press, 2018.
- Understanding Machine Learning, by Shai Shalev-Shwartz and Shai Ben-David, Cambridge University Press, 2018.
- Deep learning, by Ian Goodfellow, Yoshua Bengio, and Aaron Courville, MIT Press, 2016. https ://www.deeplearningbook.org/

References


北京大数据研究院
博雅大数据学院系列教材
数据科学导引
欧高炎
朱占星
董彬
鄂维南
高等教育出版社


周志华 著
MACHINE LEARNING
机器学习
清华大学出版社


Springer Series in Statistics
Trevor Hastie
Robert Tibshirani
Jerome Friedman
The Elements of Statistical Learning
Data Mining, Inference, and Prediction
Second Edition
Springer