Lecture 1: Course Overview & Expectations

Felix Held, Mathematical Sciences

MSA220/MVE440 Statistical Learning for Big Data

23rd March 2020

CHALMERS
UNIVERSITY OF TECHNOLOGY

UNIVERSITY OF GOTHENBURG

Course Overview & Expectations

Who's involved

Felix Held, felix.held@chalmers.se
lectures and course coordination

Rebecka Jörnsten, *examiner*

Sebastian Persson
sebpe@chalmers.se

Juan Inda Diaz
inda@chalmers.se

Think before you write

Before you send an email/write a message

- Check on Canvas (Modules, Course PM, Syllabus, ...)
- Review the lecture slides
- If it's a content question: Think a little bit about the problem first and then **post in Discussions on Canvas**.

A course in three parts

1. Lectures
2. Projects
3. Take-home exam

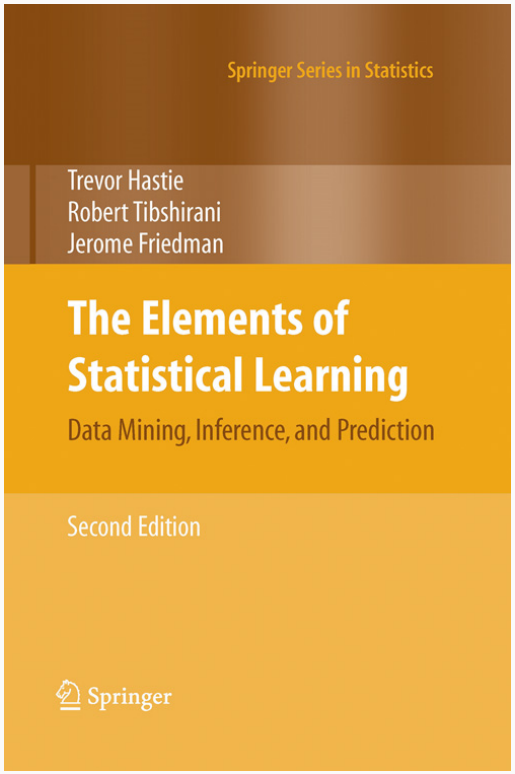
Themes

- Statistical learning/prediction: Regression and classification
- Unsupervised classification, i.e. clustering
- Variable selection, both explicit and implicit
- Data representations/Dimension reduction
- Large sample methods

Course limitations

- Understanding of **algorithms**, **modelling assumptions** and **reasonable interpretations** are our main goals.
- We will focus on well-understood and interpretable methods and their modifications for big data sets.
- This course focuses on the statistics and not on the logistics of data.
- No neural networks or deep learning. There are specialised courses for this (e.g. FFR135/FIM720 or TDA231/DIT380) and as of today, these are mostly black-box models.

Course literature



Hastie, T, Tibshirani, R, and Friedman, J (2009) The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. New York: Springer Science+Business Media, LLC

- Covers a lot of statistical methods
- Freely available online
- Balanced presentation of theory and application
- Not always very detailed. Other suggestions on course website.

Projects

- **Four small projects** throughout the course
- **Purpose:**
  - Hands-on experience in data-analysis
  - Further exploration of course topics
  - Practice how to present statistical results
- The details are still being discussed, but preliminary deadlines for hand-ins are
  - 3rd April
  - 24th April
  - 8th May
  - 15th May
- More information by the end of the week

Exam

- **Take-home exam**
- **Structure:** A couple of data analysis tasks that need to be answered in form of individual reports.
- Exam will be handed out on 28^{th} May
- Hard deadline on 12^{th} June