Lecture 1: Course Overview & Expectations

Felix Held, Mathematical Sciences

MSA220/MVE441 Statistical Learning for Big Data

22nd March 2021



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 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.

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

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

- 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.

Course literature

Trevor Hastie Robert Tibshirani Jerome Friedman The Elements of Statistical Learning Data Mining, Inference, and Prediction

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.

Three projects throughout the course

Starting from this year, projects are a separate course module in Ladok

Purpose:

- Hands-on experience in data-analysis
- Further exploration of course topics
- Practice how to present statistical results
- Last but not least, preparation for the take-home exam

Projects (II)

- ▶ Two weeks for each project, starting after Easter (week 15)
- Dates for the projects are on Canvas
- Structure:
 - Choose among a small selection of topics
 - After one week, there will be a discussion session on Zoom. The dates are 19th April, 3rd May, and 17th May.
 - > You create a short presentation summarising your results concisely
 - ▶ Juan, Sebastian, and I will give you feedback on your results
 - Insufficient submissions may need to be reworked
- Guidelines on what a sufficient submission is will be published on Canvas
- ► If possible, please work and submit in pairs

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 27th May
- ▶ Hard deadline on 11th June