Preliminary Course Outline

Instructor: Rebecka Jörnsten, `jornsten@chalmers.se`
Class homepage: `http://www.math.chalmers.se/Stat/Grundutb/GU/MSA220/S18//`
Lectures: Mon 3-5pm (Euler), Thur 10-12am (Euler), Fri 1-3pm (MVH12).
Office hours: Mon 2-3 pm, Thursday 9-10am in MVH 3029.

This is the preliminary course outline only. I will post updates on book chapters and additional materials like journal papers as the semester progresses. The outline presents the main topic for each week. There are 4 main themes and 3 big data settings we will consider:

Themes
 – Model building (regression and classification)
 – Data representations
 – Clustering (which is a form of data representation in fact)
 – Large-sample methods

Settings
 – Big-p: high-dimensional data with many variables
 – Big-n and Big-p: high-dimensional and large sample
 – Big-n: batch estimation, sub-sampling to deal with a large number of observations

I have structured the course as follows: 1st week of a theme we will look at "classical" multivariate methods, albeit some of the quite modern (and some not so much), 2nd week of a theme we will look at how these methods need to be altered to work in Big Data settings.

For the 1st week and some of the 2nd week topics, the text book will cover most of the material needed. For some of the 2nd week topics I will provide journal articles and lecture notes.

Week	Topics	Material	Project
w12	Introduction. Clustering.	2.1-2.3, 8.7, 9.2, 15.2, 14.1-14.3	Mini1, due Thur April 12
w15	Model-building: classification and regression	2.1-2.7, 3.1-3.8, 4.1-4.4, 7.1-7.10, 13.3	Mini2, due April 19
w16	Big-p: Sparse modeling	3.8, 18.2-18.4, 18.6	Mini3, due April 26
w17	Data representations.	14.4-14.9	Mini4, due May 3
w18	Big n and p: stochastic methods	Journal papers	
w19	Clustering, Big n and p	14.1-14.3 + journal papers	Mini5, due May 17
w20	Big n: D&C, BLB, Bayesian vs Frequentists	Journal papers	Mini6, due May 24

The final grade will be based on 6 Mini-Analysis projects and 1 final individual project.
Text: *2nd edition* The Elements of Statistical Learning , Hastie, T., Tibshirani, R., and Friedman, J. (see weblink on home page), journal papers and Lecture notes.

Software: R. R is free - download for windows, linux and mac available.

For help with computing, Getting Started with R and Rstudio provides everything from the basics to links to more advanced programming.
The Art of R programming and Advanced R are online books.
There's even a tutorial on YouTube!!!

Projects

You will be required to complete 6+1 projects. For the first 6 projects you can work in pairs if you want, but not the same pairs! You will prepare to present results to the class and hand in slides electronically. Due dates are always THURSDAYS and it is on THURSDAYS you will present results to the rest of the class.

Please note: all projects are compulsory and the in-class presentations are mandatory! If you don't attend the Minis you are required to submit a report instead of the presentation slides.

The final project is an individual project. The Minis count for 50% of your final grade.

Practical Stuff

1. Class email list! This is how I will communicate with you, including Doodles - make sure you are added to the list ASAP. Send an empty email to `jornsten@chalmers.se` with the topic MSA220.

2. Office hours. Mon 2-3pm and Thur 12noon-1pm in MVH3029. Come to discuss projects or general questions.