Contents

Course information

See on the web:

- the current course content, a.y. 2018/2019.
- a past course overview, a.y. 2017/2018.

Note: if you look for information of any course go to *ricerca insegnamenti* and type either thr course name or teacher name (e.g., input trevisani).

Course overview

This is an introductory-level course on **machine/statistical learning**, **data management**, **programming**, and **visualization** using **data examples**.

Students will have a basic understanding of how learning from data to find underlying patterns useful for understanding and predicting the processes under investigation.

The course aims to provide a practical survey of modern learning techniques: **regression and classification** methods, **resampling methods** and **model selection**, **regularization**, **tree-based methods**, **support vector machines**, **principal components analysis**, and **clustering methods**.

Time and place

This course meets in Room 3B of the H2 Building on Mondays, Tuesdays and Wednesdays, 9:15 am–12:30 am.

There may be exceptions to these meeting times (for other commitments or impediments of teacher, guest lectures, holidays, etc)

Detailed schedule

Here is a general scheme of the schedule for this course for Fall 2018:

chunk	CFU	hours	to	teacher	weeks
pre-course (R)	0	6	All	Pelle	Oct 1-5
T1	3	24	DSSC	Trevisani	Oct 8-27
M1	3	24	All	Medvet	Nov 5-16
T2	3	24	All	Trevisani	Nov 19-30
M2	3	24	All	Medvet	Dec 10-21

First chunk (M1, just for DSSC students):

chunk	week	content
M1	Oct 8-12	Introduction, data science toolbox, prob/stat recap
	15-19	Elements of statistical learning, regression function, model accuracy, bias-variance trade-off
	22-26	Supervised learning with linear models, model checking, model selection, qualitative predictors, regula

Grading

Course grades will be determined by:

- either a full written test
- or a short written test + student project (highly preferable to be made in groups): the vote is the average of written/application

Rules:

- you can take exam in any date and withdraw at any time, but
- when you deliver your test you cannot reject the vote.

Temptative schedule:

- 21/1, 4/2, 14/2;
- 12/6, 26/6;
- 16/9 (2/9)

Attendance and class participation

Because so much of the learning for this class is through interactive work in class, it is critical that you come to class.

Weekly in-course group exercises

Part of each class will be spent doing in-course group exercises.

Tutoring

Marco Zullich (marco.zullich.at.gmail.com)

- wed 6-7 pm (math dept) for DSSC e SSA
- tue (after class), about 1-2 pm (possibly anticipating) mainly for ING
- you may ask for an appointment
- occasional participation during lab in class

Laptops

- Please plan to bring a personal laptop to all classes.
- If you do not have access to a laptop you can bring to class, talk to me and we can try to figure something out.

Course set-up

Please be sure you have the latest version of R and RStudio (Desktop version, Open Source edition) installed. Both are free for anyone to download.

Here are useful links for this set-up:

- R: https://cran.r-project.org
- RStudio: https://www.rstudio.com/products/rstudio/#Desktop

Suggested textbooks

- Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani. An Introduction to Statistical Learning, with applications in R. Springer, Berlin: Springer Series in Statistics, 2014.
- Jerome Friedman, Trevor Hastie, and Robert Tibshirani. The elements of statistical learning: Data Mining, Inference, and Prediction. Springer, Berlin: Springer Series in Statistics, 2009.
- Kenneth A. De Jong. Evolutionary computation: a unified approach. MIT press, 2006.

Other helpful books (not required)

- Elements of Data Analytic Style
- The Art of Data Science
- R for Data Science
- Exploratory Data Analysis
- Regression Modeling for Data Science in R

There are a number of other useful books available on general R programming, including:

- R for Dummies
- R Cookbook
- R Graphics Cookbook
- Roger Peng's Leanpub books

The R programming language is used extensively within certain fields, including statistics and bioinformatics. If you are using R for a specific type of analysis, you will be able to find many books with advice on using R for both general and specific statistical analysis.