An introduction to Statistical Learning for Atmospheric and Climate Sciences

28-29 August 2018

A two-day block course on "An introduction to Statistical Learning for Atmospheric and Climate Sciences" will be held at the Institute for Atmospheric and Climate Sciences, ETH Zurich on 28-29 August 2018. The course will provide an introduction to Statistical Learning methods for beginners and practitioners in the context of Atmospheric and Climate Science applications, featuring both overview lectures of applications and methods, and hands-on exercises.

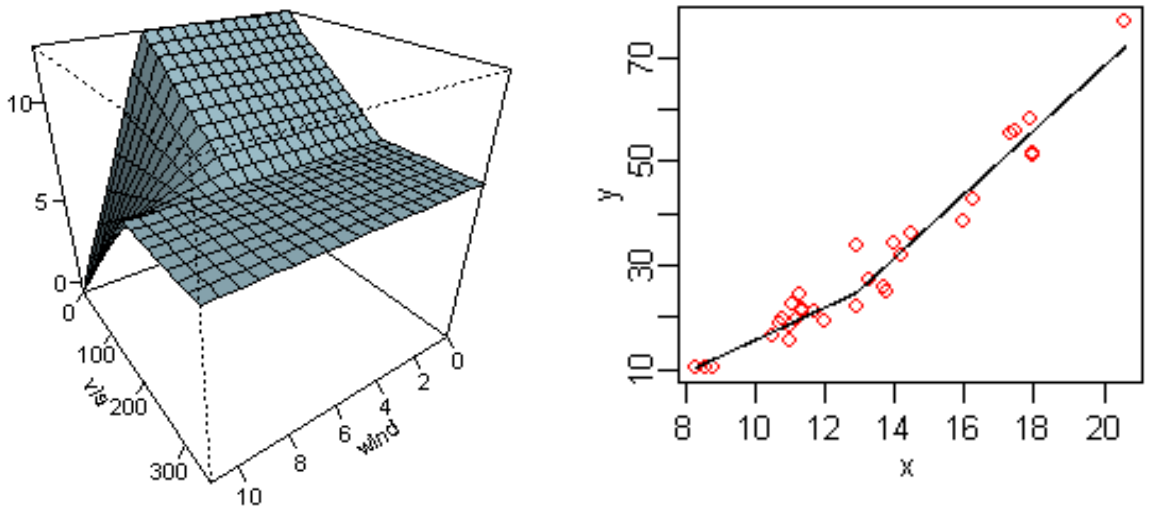

Image source: https://en.wikipedia.org/wiki/Multivariate_adaptive_regression_splines

Target group
The anticipated target group for the workshop consists of M.Sc/PhD students and PostDocs, who have a (practical) interest in Statistical Learning / Machine Learning / Data Science, but who had so far only little or no experience to Statistical Learning, and thus would like to gain an introductory overview. The course is open to everyone interested, exercises are limited to 30 participants (on a first-come-first-serve basis). No credit points can be given for the course.

Content & Lecturers
The course will consist of alternating overview lectures and hands-on practical exercises. Lectures will cover theoretical basics of Statistical Learning (Exploratory Analysis, Linear regression, sparse linear models, non-linear supervised learning) and a general overview and outro of applications of Statistical Learning in the Atmospheric and Climate Sciences. Exercises will allow the participants to understand, test and evaluate different methods in a "real-world" climate-science example (for details, see program attached). Exercises will be done in the R environment (https://www.r-project.org/), which is a specialized tool for statistical computing.

Venue
Lectures and exercises take place in the main building of ETH Zurich: Lectures: HG D 3.2; Exercises: HG E 19.

Costs & Registration
The course is free. Please register through following web page:
http://www.iac.ethz.ch/edu/technical-training/statistical-learning-atm-clim-sci.html

Note that the number of participants for the Exercises is limited to 30.

Short Program:

Day 1 morning

General Introduction
- Why Statistical learning in climate science?
- Conceptual overview of statistical learning applications in climate science

Theory 1
- Bias-Variance trade-off
- Simple & Advanced linear regression
- Cross-validation & hyperparameter selection

Day 1 afternoon

Exercises 1 & 2
- Bias-Variance trade-off
- Simple & Advanced linear regression

Day 2 morning

Theory 2 & Exercise 3
- Nonparametric supervised learning (random forests, bagging, neural networks)

Day 2 afternoon

Exercise 4
- Statistical Learning Challenge

Nonlinear Dimension Reduction
Outro: Advanced Machine Learning topics

Detailed Program:

	Day 1: 28.08.2018	Day 2: 29.08.2018
Morning 1: **09.00 – 10.30**	**Introduction** (**HG D 3.2**, SISe/RKn):Why Statistical Learning in climate science Emergence of "Big Data" Brief conceptual overview of ML use cases in climate science	**Theory 2** (**HG D 3.2**, SSi/LGu): Nonparametric supervised learning (Random Forests, Bagging, Neural Networks)
Coffee Break		
Morning 2: **11.00 – 12:30**	**Theory 1** (**HG D 3.2**, SSi/LGu): Bias-Variance Trade-off Cross validation Advanced linear regression (Ridge, LASSO)	**Exercise 3** (**HG E 19**, SSi/LGu): Nonparametric supervised learning
Lunch		
Afternoon 1: **13.30 – 15.00**	**Exercise 1** (**HG E 19**, SSi/LGu): Bias-variance trade-off Cross validation	**Exercise 4** (**HG E 19**, SSi/LGu): Statistical Learning challenge
Coffee Break		
Afternoon 2: **15.30 – 16.45**	**Exercise 2** (**HG E 19**, SSi/LGu): Advanced linear regression (Ridge, LASSO)	**Nonlinear Dimension Reduction** (**HG D 3.2**, ESz) **Advanced ML topics** (**HG D 3.2**, NMe): *Outro:* what are emerging topics in ML that are relevant for climate science?
16.45-17.00	**Short Feedback / Wrap-up**	**Feedback/Discussion/Wrap-Up**

Names: SiSe: Sonia I. Seneviratne; RKn: Reto Knutti; NMe: Nicolai Meinshausen; SSi: Sebastian Sippel; LGu: Lukas Gudmundsson; ESz: Eniko Szekely

Literature (on which the course will be largely based):
Hastie, T., Tibshirani, R. & Friedman, J. (2009). The elements of statistical learning (Vol. 2). New York: Springer series in statistics.
(Link to book: https://web.stanford.edu/~hastie/Papers/ESLII.pdf, book homepage: http://web.stanford.edu/~hastie/ElemStatLearn/)

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112). New York: springer.
(Link to book: http://www-bcf.usc.edu/~gareth/ISL/ISLR%20First%20Printing.pdf, book homepage (exercises, etc.): http://www-bcf.usc.edu/~gareth/ISL/)