Machine Learning

Course outline

JS

August, 2022

Course title	:	Machine learning
Lecturer	:	Jonas Striaukas. Email: js.fi@cbs.dk or jonas.striaukas@gmail.com
Lecture time	:	TBA
Auditorium	:	TBA
Course website	:	https://jstriaukas.github.io/ml_course

Important : please write *Machine learning class* in the subject line when you write to me regarding the course material/questions/etc. Please write questions after each lecture, I will allocate 10-15 minutes at the beginning of the following lecture to answer your questions.

Prerequisites : introduction to statistics, linear regression and some basic computing in statistical software (e.g., R, Python) is assumed, but otherwise it is a self-contained course. I recommend to read through the introductory book (first book in the list of recommended books) before the course to have a rough idea on what we will cover during the course.

Books : I try to make material self-contained so that you don't need to buy any book for the course. However, if you like the course and want to have a deeper understanding on a particular topic or machine learning in general, I suggest the following books:[1]

- (great introductory book) James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, p. 18). Springer (New York).
  ▶ online copy: pdf.
- (moderate level overview book) Hastie, T., & Friedman, J. H., Tibshirani, R. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2, pp. 1-758). Springer (New York).
  ▶ online copy: pdf.
- (main book for the course – covers more recent topics) Fan, J., Li, R., Zhang, C. H., & Zou, H. (2020). *Statistical foundations of data science*. Chapman and Hall/CRC.

Software: students need to install statistical software R or some other program (Python, Matlab, Julia, Octave, ...). I will be showing examples mainly in R and Python. I advise you to install the statistical software before the course starts (if you don't have it already). See below for the instructions how to install R and Python. You are free to use Matlab or Julia or other language of your choice to code the examples we cover in the class.

Exam: TBA

[1] I list additional set of books for students who want to learn more about statistical learning methods, statistical theory, introductory probability theory for high-dimensional problems, etc., at the end of this outline.

Information about the course

Topics

The course will cover the following main topics:

- **Topic 1:** Introduction to learning, multiple and nonparametric regression
  ▶ Material: slides shinyapps slides pdf tablet friendly slides pdf
- **Topic 2:** High-dimensional linear regression
  ▶ Material: slides shinyapps slides pdf tablet friendly slides pdf
- **Topic 3:** High-dimensional regression properties and generalized linear models (GAMs)
  ▶ Material: slides shinyapps slides pdf tablet friendly slides pdf
- **Topic 4:** Prediction, loss functions and M-estimators
  ▶ Material: slides shinyapps slides pdf tablet friendly slides pdf
- **Topic 5:** Introduction to deep learning
  ▶ Material: slides shinyapps slides pdf tablet friendly slides pdf
- **Topic 6:** Introduction to causal machine learning
  ▶ Material: slides shinyapps slides pdf tablet friendly slides pdf

Details on each topic

Topic 1: *Introduction to learning, multiple and nonparametric regression.*
The section starts with reviewing challenges and advantages of *Big data* and a brief introduction to learning (theory).[2] The section then covers multiple and nonparametric regression methods which are central to learning. Specific examples of the techniques that ... .

Instructions to install R

The main R software is available at https://cran.r-project.org. Once on this website, you need to select an file to download for your operating system (OS), so if you work on Windows you need to download and install Window `.exe` file. Please install the most recent version of R. I also strongly advise to install RStudio (free version of it) from https://rstudio.com. You need to install R prior to RStudio.

Install R packages: suppose you need to install an R package called '`forecast`'. You should write in RStudio console:

```
install.packages("forecast")
```

[2] **IMPORTANT:** I write theory in brackets as throughout the course theoretical derivations serve us the purpose of understanding certain learning technique, how and why it works for certain data types. I do not expect that you reproduce more complicated proofs, however, I expect that you understand them and can interpret results, i.e. can argue why certain methods work for certain data types.

List of useful books

- (introductory book on high-dimensional stats) Giraud, C. (2021). *UItroduction to High-Dimensional Statistics* (Vol. 112, p. 18). Chapman & Hall/CRC Monographs on Statistics and Applied Probability.
- (great book for LASSO methods) Bühlmann, P., & Van De Geer, S. (2011). *Statistics for high-dimensional data: Methods, theory and applications.* Springer Science & Business Media.
- (more advanced book on high-dimensional stats) Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint* (Vol. 48). Cambridge University Press.
- (introductory probability theory for high-dimensional problems) Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science* (Vol. 47). Cambridge university press.
- (reference book for concentration inequalities) Boucheron, S., Lugosi, G., & Massart, P. (2013). *Concentration inequalities: A non-asymptotic theory of independence*. Oxford university press.