Machine learning and spatial analysis

Syllabus

Clément GORIN (CNRS) gorin@gate.cnrs.fr

1. Overview

This course introduces a range of modern research techniques to deal with high dimensional and potentially unstructured data. These techniques fall into two distinct categories. The first part of the course introduces statistical learning techniques. In recent years, there has been an increasing interest in predictive modelling, in particular for the collection of original research data that would not be available using traditional methods. This course provides a comprehensive understanding of some of the most capable supervised learning algorithms, including GAMs, random forests, boosted trees, or neural networks. The second part of the course focuses on applied spatial economics. Many economic phenomena are spatial in nature. Manipulating and analysing spatial data rely on specific sets of tools known as geographical information systems. Specifically, students will learn how to manage vector and raster data, perform geocomputations, represent spatial processes and fit spatial models. This course contains comprehensive theorising and mathematical formalisation but keeps a strong focus on intuition and effective implementation. In particular, we make extensive use the R programming language, both to illustrate abstract statistical concepts using simulated data, and to perform economic analysis on actual datasets.

2. Course content

The course is organised in four three-hours sessions. Each session articulates a theoretical lecture and practical applications using the R language. The sessions are organised as follows:

Session	Topic	Duration
1	Statistical learning	3 hours
2	Neural networks	3 hours
3	Trees, bagging and boosting	3 hours
4	Spatial analysis	3 hours

All lectures, exercises, solutions and resources will be made available on the Dropbox of the course. You should come to class with your computer with administrator rights, as well as a working internet connection. Make sure to install the latest stable release of R, available on the CRAN website for all platforms. In addition, you may install an IDE / GUI such as RStudio, also available for all platforms on their website.

3. Skills and requirements

This course helps you develop a solid theoretical background in machine learning, as well as the ability to understand and implement a variety of machine learning models for economic analysis. Besides, you will acquire a good understanding of spatial data formats and spatial computations, that will enable you to analyse spatial economic phenomenon. Last but not least, you will achieve greater proficiency in the R language, which enable you to handle data and estimate models with great efficiency and flexibility.

This course requires a good background in econometrics and statistics. Besides, a good understanding of an abstract programming language such as R or Python is recommended.

4. References

Papers

- Athey, Susan R. The impact of machine learning on economics. National Bureau of Economic Research, 2018.
- LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. Deep learning. Nature, 521: 436-444, 2015.
- Natekin, Alexey and Alois Knoll. Gradient boosting machines, a tutorial. Frontiers in neurorobotics, 7(21), 2013.
- Sendhil, Mullainathan and Spiess Jann. Machine learning: An applied econometric approach. Journal of Economic Perspective, 31(2): 87-106, 2017.
- Varian Hal R. Big data: New tricks for econometrics. Journal of Economic Perspective, 28(2): 3-28, 2014.

Additional references will be given in class.

Manuals

- Gareth, James, Trevor Hastie, Robert Tibshirani, and Daniela Witten. An introduction to statistical learning with applications in R. Springer-Verlag, 2014.
- Bivand, Roger S, Edzer Pebesma, and Virgilio Gomez-Rubio. Applied spatial data analysis with R. Springer-Verlag, 2013.
- Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep learning. MIT Press, 2016.
- Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. The elements of statistical learning. Springer, 2009.
- Pebesma Edzer and Roger Bivand. Spatial Data Science. 2020.
- Wickham Hadley. Advanced R. Chapman and Hall, 2019.