PhD Course 2022-2023

Machine learning and spatial analysis

Clément Gorin
Paris School of Economics
gorinclem@gmail.com

Course's objective

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 spatial computations, 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.

Outline

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:

Lecture 1: Statistical learning
Function approximation – Parameter estimation – Prediction and inference – Regularisation – Bias and variance – Resampling – Splines – Generalised additive models

Lecture 2: Neural networks
Units and layers – Activation – Forward propagation – Optimisation – Backpropgation – Optimisation issues – Convolutional networks – Application

Lecture 3: Trees, bagging, boosting

Prediction trees – Optimisation – Regularisation – Bootstrap aggregation – Random forest – Gradient boosted trees – Application

Lecture 4: Spatial analysis
Vector data – Raster data – Projections – Geographical information systems – Mapping – Spatial computations – Spatial modelling – Application

Course materials

All lectures, exercises, solutions and resources will be made available on the Dropbox of the course. Students should come to class with their computer with administrator rights, as well as a working internet connection. The practical exercises require a recent release of R, available on the CRAN website for all platforms.

Grading

Students will be graded on the basis of a final assessment. They will have to hand in a short report (10 pages maximum) along with a script reproducing the results. The report must use of machine learning and/or spatial analysis tools (possibly with econometrics) to solve a research question of interest to economists. Ideally, students should work on a topic related to their thesis.

Reference handbooks

- Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning*. Springer, 2009.
- Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT Press, 2016.

Complementary handbooks

- 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.
- Pebesma Edzer and Roger Bivand. *Spatial Data Science*. 2020.
- Wickham Hadley. *Advanced R*. Chapman and Hall, 2019.

Specific references are available on request.

Other references

- 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.