Machine Learning for Social Sciences

Jorge Cimentada

19 November, 2019

Course description

With increasing amounts of data being collected on a daily basis, Machine Learning algorithms have gained traction in the statistical world. Machine Learning techniques, which shift from understand statistical relationships to maximizing predictive power,

Machine Learning (ML) is a technique that has been gaining traction in recent years. Historically, most of it's application have been limited to the field of computer science but recently there's been ground breaking applications in the fields such as business, biology, medicine and in the social sciences. In particular, the advent of Machine Learning into the social sciences implies a shift in statistical reasoning from explanation to prediction.

Schedule

Session 1
July 6th 09h-10:45h

- Introduction to the Machine Learning Framework
  - Inference vs Prediction
  - Can inference and prediction complement each other?
  - "The Fragile Families Challenge"
  - Bias-variance / Interpretability-prediction tradeoffs
  - Resampling methods: validation, k-fold CV

Break 10:45h-11:15h

Session 2
July 6th 11:15h-13:00h

- Linear regression and reguarlization
  - Continuous predictions and loss functions
  - Lasso
    - Testing

Software:

We will be using the R software together with the Rstudio interface. No laptop is required as the seminars will take place in the RECSM facilities. Any packages we plan to use will be already downloaded previous to the session.

Prerequisites:

- The course assumes that the student is familiar with R and should be familiar with reading, manipulating and cleaning data frames. Ideally, the student has conducted some type of research using the software.
- Students should have solid knowledge of basic statistics such as linear and logistic regression, ideally with more advanced concepts such as multilevel modelling.

Instructor:

Jorge Cimentada has a PhD in Sociology from Pompeu Fabra University and is currently a Research Scientist at the Laboratory of Digital and Computational Demography at the Max Planck Institute for Demographic Resarch. His research is mainly focused on the study of educational inequality, inequality in spatial mobility and computational social science. He has worked on data science projects both in the private sector and in academic research and is interested in merging cuttind edge machine learning techniques with classical social statistics. You can check out his blog at cimentadaj.github.io or contact him through twitter at @cimentadaj.