

Topic course – Prof. Paolo Brunori

Introduction to machine learning using R

Machine learning is transforming the way in which economists think and use statistics to do research. Algorithms used by machine learners are not new, they have been used in informatics and engineering for decades. What is happening today is that the explosion of available information and the improvement in computational power of machines are making these algorithms more powerful relative to traditional econometric tools.

With computers becoming more and more involved in economic transactions, this trend is likely to be reinforced in the next future. This is why it is a good idea to acquire some rudiments about how machine learning can be used to approach econometric issues.

During the tutorial, basic concepts of machine learning and a few data manipulation tools will be introduced using real data and applications using the open source software R.

The tutorial covers three topics:

1. **K-fold cross validation**: understand the workhorse of machine learning;
2. **Regression trees and random forest**: a basic tool for visualizing data and predict out-of-sample;
3. **Regularization (LASSO, ridge regression, elastic net)**: implement penalized regression in R to prevent model overfitting.

Software used: R (<http://www.r-project.org/>)

If you want to be able to replicate the code during the tutorial, please Install the last version of R and R Studio on your laptop.

Knowledge required: undergraduate level of statistics and econometrics.

Reading list:

- Hal Varian. 2014. “*Big Data: New Tricks for Econometrics*”, *Journal of Economic Perspectives*, Volume 28, Number 2.
- Selected section of: James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. “*An Introduction to Statistical Learning: With Applications in R*”. New York: Springer.
-

Lecturer: Paolo Brunori, *University of Florence & London School of Economics*