Applied Machine Learning with R

Cleyton Farias

March 27-29, 2019

E-mail: cleytonfarias@outlook.com
Web: cleytonfar.github.io
Office Hours: TBD
Class Hours: 13h-18h
Office: TBD
Class Room: 209 CCSA

Course Description

In recent years, data science has become a huge attractive field with the profession of data scientist topping the list of the best jobs in America. And with all the hype that the field produces, one might ask: what does it take do be a data scientist?

In this course you will get an introduction to the fascinating world of working with data and extract something valuable from it. You will learn what are the main tools in the data scientist's toolbox. That is, you will learn how to perform a predictive analysis with machine learning algorithms using real world applications. Step by step, we are going to build models that can be used to predict either categorical and continuous variables.

Differently from most available courses, this one will take one more step forward and show you how to improve your models in order to get better results. Moreover, you will be exposed to some applications of machine learning in economics problems.

Course Objectives

1. Learn what are the main tools in the data scientist's toolbox;
2. Introduce to the world of machine learning and predictive analysis;
3. Develop machine learning model, evaluate and improve them;
4. Machine learning applications in economic problems;

Syllabus:

1. Introduction to Predictive Analysis:
   - What is predictive analysis;
   - Bias-variance trade-off;
   - Supervised vs Unsupervised Learning;
   - Regression vs Classification.
2. Algorithms:
   - Penalized methods: Ridge, LASSO, Elastic net;
   - Tree-based Models: Decision Trees, Bagging, Random Forest, Boosting;
   - Support Vector Machines;

   - Neural Networks.
3. Evaluating Model Performance:
   - Measuring performance: accuracy, precision, recall, F-measure, kappa, ROC curve, logLoss;
   - Estimating future performance: holdout, cross-validation methods.
4. Economics applications:
   - Source of Economic Growth: Barro and Lee (1994) revisited;
   - Can Machine Learning help predict the outcome of asylum adjudication? Chen(2019);
   - Estimating the Effect of Institutions on Output: Acemoglu (2001) revisited;

Requirements:

This course will be taught using the **R** programming language. Also, some packages will be necessary throughout this course. These packages are:

- *ISLR*;
- *MASS*;
- *class*;
- *tree*;
- *randomForest*;
- *gbm*;
- *glmnet*;
- *e1071*;
- *caret*;
- *keras*;

Some references fo the course include:

Chen, Daniel L and Jess Eagel "Can machine learning help predict the outcome of asylum adjudications?" In: *Computational Analysis of Law*.

Chollet, Francois (2018). *Deep learning with R*. Manning Publication.

James, Gareth, Daniela Witten, Trevor Hastie and Robert Tibshirani (2013). *An Introduction to Statistical Learning*. Vol. 112. Springer.

Lantz, Brett (2013). *Machine learning with R*. Packt Publishing Ltd.

Hansen, Stephen, Michael McMahon and Andrea Prat (2017). "Transparency and deliberation within the FOMC: a computational linguistics approach". In: *The Quarterly Journal of Economics* 133.2, pp. 801–870.

Jean, Neal, Marshall Burke, Michael Xie, W Matthew Davis, David B Lobell and Stefano Ermon (2016). "Combining satellite imagery and machine learning to predict poverty". In: *Science* 353.6301, pp. 790–794.

Belloni, Alexandre, Victor Chernozhukov and Christian Hansen (2014). "Inference on treatment effects after selection among high-dimensional controls". In: *The Review of Economic Studies* 81.2, pp. 608–650.

Acemoglu, Daron, Simon Johnson and James A Robinson (2001). "The colonial origins of comparative development: An empirical investigation". In: *American economic review* 91.5, pp. 1369–1401.

Gentzkow, Matthew and Jesse M Shapiro (2010). "What drives media slant? Evidence from US daily newspapers". In: *Econometrica* 78.1, pp. 35–71.

Barro, Robert J and Jong-Wha Lee (1994). "Sources of economic growth". In: *Carnegie-Rochester conference series on public policy*. Ed. by dea. Vol. 40. Elsevier. , pp. 1–46.