CART and Random Forests

Jean-Michel Poggi

Univ. Paris & LMO, Univ. Paris-Saclay, Orsay

Overview of the Course

Universidad de la República – Facultad de Ingeniería

Instituto de Matemática y Estadística "Prof. Ing. Rafael Laguardia"

Montevideo, Uruguay

February, 10-14, 2020

Overview

- ***CART trees***
- *Spatial CART*

- ***Random Forests, Variable importance and Variable Selection***
- *An application to Driver's Stress Level Classification*
- ***Random Forests for Big Data***

Course 1 - *CART trees*

- 1. Introduction
- 2. CART trees and splits
- 3. Construction: maximal tree and pruning
- 4. A typical theoretical result from the model selection viewpoint
- 5. Extensions and variants
- 6. CART in practice using the R package `rpart`

- Breiman, L., Friedman, J., Olshen, R. et Stone, C. [1984]. *Classification and Regression Trees* . Chapman & Hall, New York.
- Therneau, T., Atkinson, B. et Ripley, B. [2015]. `rpart`: *Recursive Partitioning and Regression Trees*

Course 1 (continuing)
Spatial CART

- 1. Spatial CART
- 2. Influence Measures for CART
- 3. Influence Measures and Stability for Graphical Models

- Bar-Hen, A., Gey, S. et Poggi, J.-M. [2019]. *Spatial CART classification trees*. Submitted.
- Bar-Hen, A., Gey, S. et Poggi, J.-M. [2015]. *Influence measures for CART classification trees*. Journal of Classification, 32 (1), 21–45.
- A. Bar-Hen, J-M. Poggi [2016] *Influence Measures and Stability for Graphical Models*, Journal of Multivariate Analysis, Vol. 47, 145-154

Course 2 *Random Forests, Variable importance and Variable Selection*

- 1. Introduction
- 2. Trees, Bagging and Random Forests (RF)
- 3. Extensions and variants, some theoretical results
- 4. Out-of-bag error and variable importance measure
- 5. Variable Selection using RF
- 6. RF in practice using the R packages `randomForest` and `VSURF`

- Breiman, L. [2001]. *Random forests*. Machine learning , 45 (1), 5–32.
- Genuer, R., Poggi, J.-M. et Tuleau-Malot, C. [2010]. *Variable selection using random forests*. Pattern Recognition Letters , 31 (14), 2225–2236.
- Genuer, R., Poggi, J.-M. et Tuleau-Malot, C. [2015]. `Vsurf`: *An R package for variable selection using random forests*. The R Journal , 7 (2), 19–33.

Course 3 (first part)
An application: RF-Based Approach for Physiological Functional Variable Selection

- 1. Introduction and motivation
- 2. Physiological functional variables and wavelets
- 3. Block variables importance measure
- 4. Functional variable selection using RF
- 5. Driver's Stress Level Classification

- Gregorutti, B., Michel, B. et Saint-Pierre, P. [2015]. *Grouped variable importance with random forests and application to multiple functional data analysis*. Computational Statistics & Data Analysis, 90, 15–35.
- El Haouij N., Poggi J.-M., Ghozi R., Sevestre Ghalila S., Jaïdane M [2017], *Random Forest-Based Approach for Physiological Functional Variable Selection: Towards Driver's Stress Level Classification*. To appear in Statistical Methods & Applications Journal.

Course 3 (second part)
Random Forests for Big Data

- 1. Motivation: Statistics in the Big Data World
- 2. Big Data characteristics and Strategies for scaling to Big Data
- 3. The reference scheme: standard RF
- 4. Sequential and parallel implementations of standard RF
- 5. m-out-of-n RF, Bag of Little Bootstraps RF, Divide-and-conquer RF, Online RF
- 6. Experimental results

- Genuer, R., Poggi, J.-M., Tuleau-Malot, C. et Villa-Vialaneix, N. [2017]. *Random Forests for Big Data*. Big Data Research, 9, 2017, 28-46 ArXiv e-prints, HAL, hal.archives-ouvertes.fr/hal-01233923v1

A final reference

- A reference freely available, in **French**, including an extensive list of references:

  Robin Genuer, Jean-Michel Poggi

  Arbres CART et Forêts aléatoires, Importance et sélection de variables

  45 pages, 2017

  https ://hal-descartes.archives-ouvertes.fr/hal-01387654v2

- published, related to the JES 2016, in:

  Maumy-Bertrand M., Saporta G. & Thomas Agnan C. (eds)

  Apprentissage Statistique et Données Massives,

  chapter 8, 295-342, 2018, Technip, Paris, France