Statistical Learning

Course Presentation

Alex Sánchez and Pedro Delicado

Universitat de Barcelona and Universitat Politécnica de Catalunya

Outline

1. Course Objectives
2. Teachers
3. Sketch of contents
4. Methodology
5. Evaluation
6. References

Course Objectives

1 - **Understand the Statistical/Machine Learning approach** and how it relates with traditional Statistical Modelling.

2 - **Develop understanding of key Learning methods**, including:

- Building and tuning models
- Appropriate application of models using relevant software tools
- Interpreting results and recognizing limitations

3 - **Adapt to Rapid Field Advancements**, staying informed about the fast-paced developments in the field and learn to manage and integrate new advancements.

Teachers (1) Alex Sanchez

Statistics and Bioinformatics

Integrative analysis of omics data

Alex Sánchez-Pla

Full Professor of Statistics.

Faculty of Biology Universitat de Barcelona

UB Director MSc of Statistics & Bioinformatics

Nutrition and Metabolomics

Former Head of Statistics and Bioinformatics Unit

Head of *Data Science Projects* Platform

Vall d'Hebron Institut de Recerca

EIT Health is supported by the EIT, a body of the European Union

Diet, Gut & Metabolomics

Software development

Teachers (2) Pedro Delicado

Pedro Delicado

Share:

(Menorca, 2015)

Full Professor

Campus Nord UPC
C5 building, level 2, room 211
C/ Jordi Girona 1-3, 08034 Barcelona

Phone: +34 93 4015698
e-mail: pedro.delicado(at)upc.edu

Member of the research group ADBD

Main interests:

Statistics and Data Science interrelation, Big Data, functional data analysis, nonparametric statistics, distance based statistical methods, multivariate data analysis, dimensionality reduction, spatial statistics, data visualization, applications of Statistics to real problems.

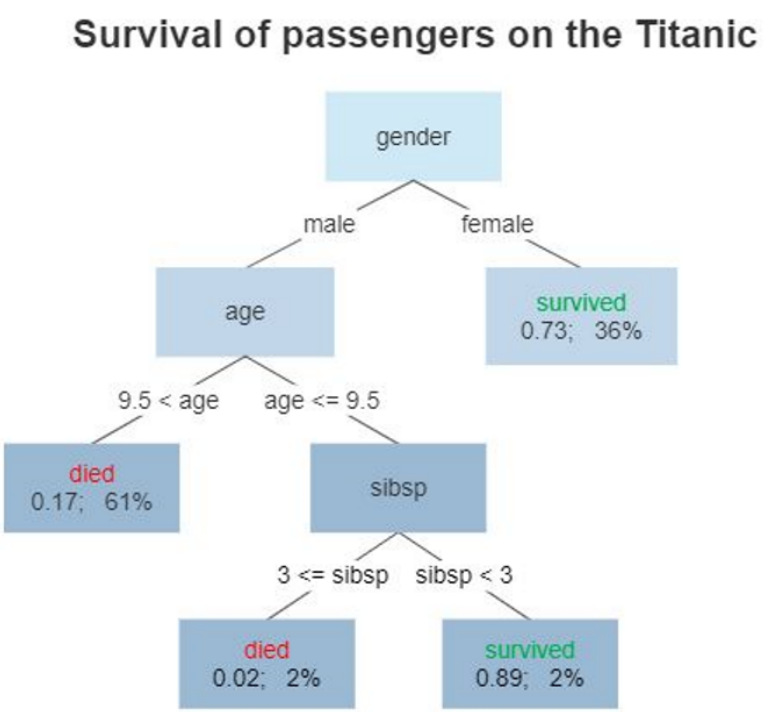
More information:

- Curriculum

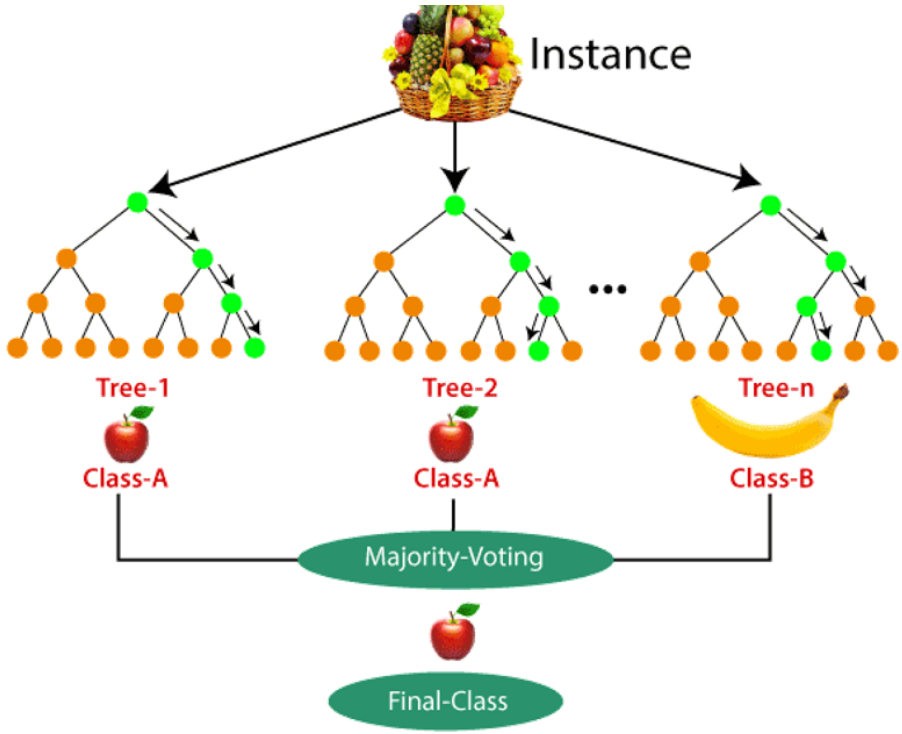
Course Contents

1. Introduction to Statistical Learning
2. Tree-based methods
3. Neural Networks
4. Penalized regression: Ridge & Lasso
5. Non parametric regression and Generalized Additive Models

Contents (1): Trees based methods

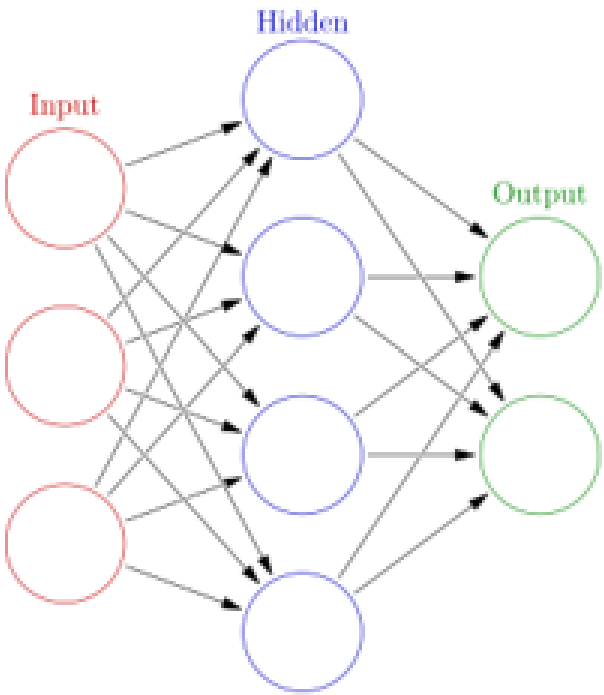


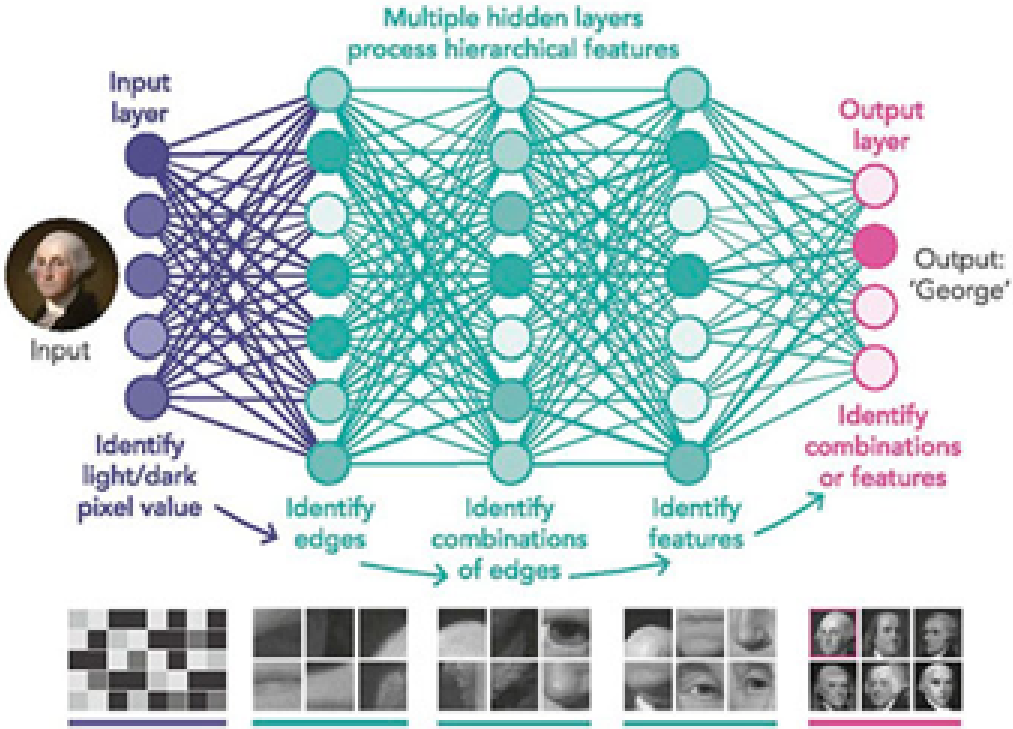
Decision Trees



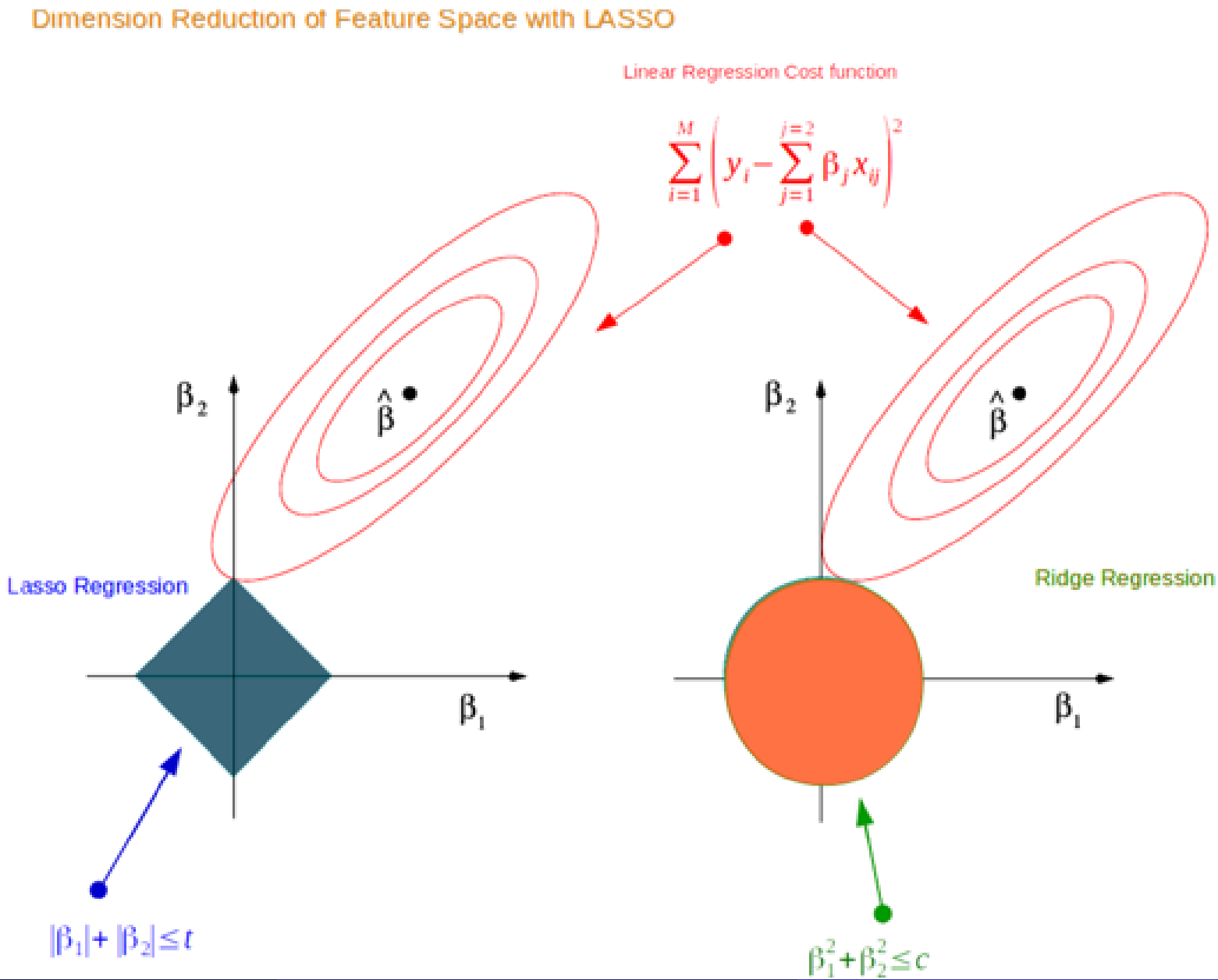
Ensemble of Trees (Random Forest)

Contents (2): Neural Networks

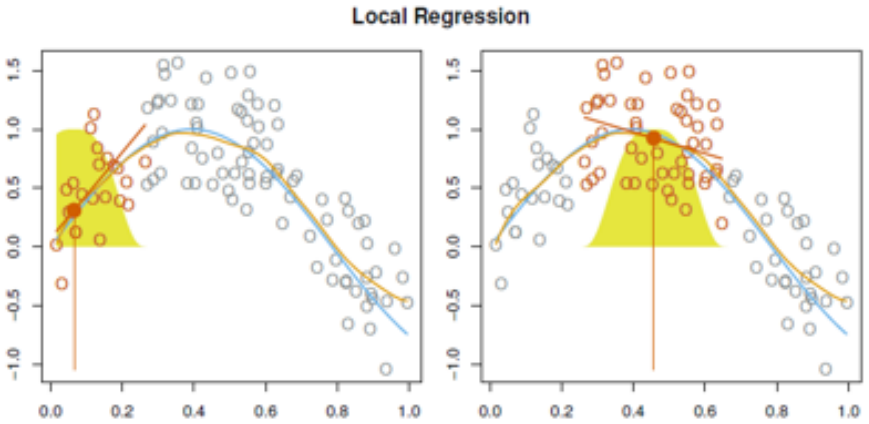


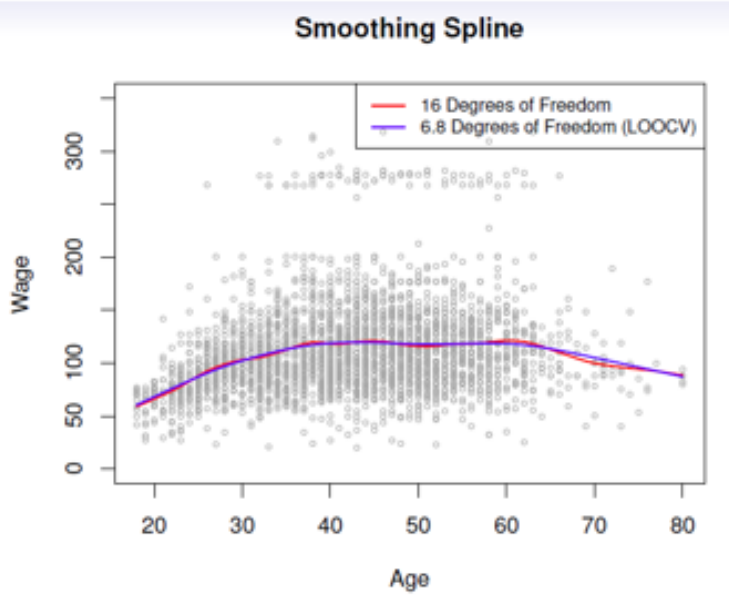


Contents (3): Penalized regressions



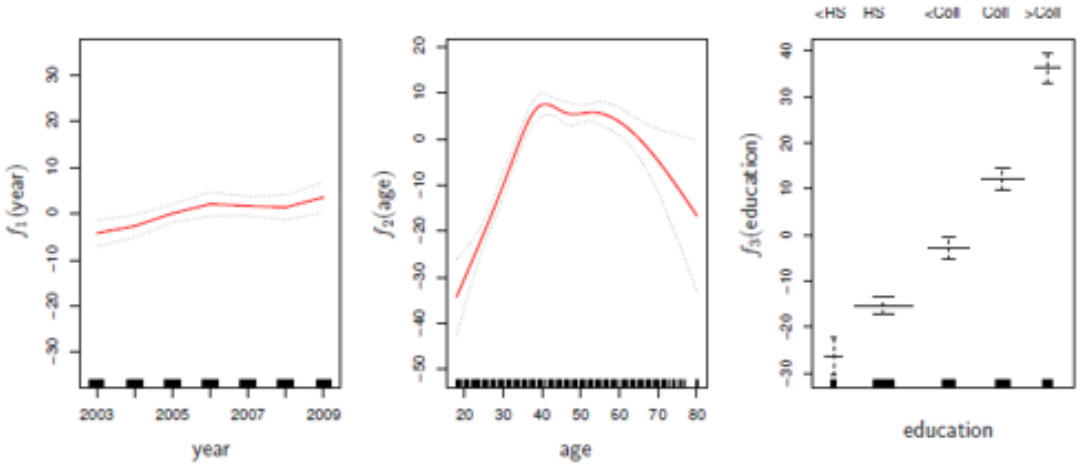
Contents (4): Non parametrics regression and GAMs





Generalized Additive Models (GAMs)

$$y_i = \beta_0 + f_1(x_{i1}) + f_2(x_{i2}) + \cdots + f_p(x_{ip}) + \epsilon_i.$$



Methodology

- Teaching language is Spanish. Materials in English.
- Main concepts will be presented in class based on slides/blackboard.
  - Materials in Atenea and/or github
- Practical applications will be demonstrated/followed using notebooks provided in campus.
  - Will try to use both R and Python
  - You are expected to be able to re-run and understand them.
- Exercises for practice will be provided and their solution will be either discussed in class or provided.
- Student participation is encouraged, either by presenting their work in class and/or contributing to the forum.

Timing

- Part 1. Alex. From 02/05/2025 till 03/26/2025
  - Introduction
  - Tree based methods
  - Neural networks
- Part 2. Pedro. From 03/31/2025 till 05/14/2025
  - Penalized regression methods
  - Non parametric regression

Evaluation & Grading

- As indicated in the course guide
- Each part of the course: 50%.
- For each part:
  - A final examen is done with weight of 50%
  - Remaining 50% is the average of scoring of submitted tasks.
- In the age of ChatGPT checking individual work is hard so there will also be on-class quizzes and oral verification of submitted work.

References and resources

- Baumer, B. S., D. T. Kaplan, and N. J. Horton (2017). Modern Data Science with R. CRC Press and Online
- Bradley Boehmke & Brandon Greenwell Hands-On Machine Learning with R, CRC press and Online
- Chollet, F. (2023). Deep learning with R . 2nd edition. Manning Publications. (*There is also a Python version*)
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning (2nd Ed). Springer.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning. Springer. Web site
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning (Vol. 1). MIT press. Web site