Machine Learning

FIB, Master in Data Science

Marta Arias, Computer Science @ UPC

Course information

Instructors

Marta Arias (theory + lab group 13)

- marias@cs.upc.edu
- no office hours; if you want to talk, email me and we'll set a time

Ignasi Gomez (lab group 11+12)

- ignasi.gomez@upc.edu

Carlos Escolano (lab group 14)

- carlos.escolano@upc.edu

Class logistics

- Course material (lecture slides, lab notebooks, project info) placed weekly through the racó
- Announcements and submissions through the racó
- Theory lectures (Mondays 2-4pm)
  - mostly presentations with slides, with occasional "old-school" chalk-board class
  - lecture notes/slides posted as we progress through the course
- Lab sessions (Tuesdays 2-4pm and 4-6pm)
  - will use `python notebooks` mostly
  - notebooks will implement and illustrate concepts from theory and introduce libraries as well
  - maybe use labs to do some examples in depth
  - may propose exercises and examples to solve *in your own time*, not graded
  - may introduce new complementary topics with examples to what is presented in theory classes

Evaluation

- Grade = 20% **mid-term** + 40% **exam** + 40% **project**

Course project

- Done in **pairs** within same lab group (this is *strict*)
- Deep pre-processing and analysis of dataset of your choice (with some limits); check with your lab tutor to see validity of your dataset when the time comes
- Intended to start early and grow mature over time
- A final **written report** (along with the code) should be carefully prepared

Delivery of project report and code towards the **end of June**; exact date will be announced in the coming weeks

Mission statement

The aim of this course is to introduce you to important ***concepts*** *in machine learning and some key machine learning methods; it is not intended to cover the latests developments in the area (which come every second) but rather to give you a solid basis that will allow you to understand new developments in the field.*

Contents[1]

Supervised learning:

- Linear methods:
  - Linear methods for regression
  - Linear methods for classification

- Non-linear methods:
  - Kernel methods (support vector machines)
  - Artificial Neural Networks
  - Random Forests and other ensemble methods

Unsupervised learning:

- Dimensionality reduction
- Clustering

[1] *Disclaimer: the topics and/or their order may change (slightly)*

Main bibliography

- Pattern Recognition and Machine Learning
  Christopher M. Bishop, Springer, 2006
- The Elements of Statistical Learning
  Hastie, Tibshirani and Friedman (2009). Springer-Verlag.
- Machine Learning: a Probabilistic Perspective
  Kevin P. Murphy, MIT Press 2012; new edition drafts here
- Introduction to Machine Learning
  Ethem Alpaydin (3rd Ed.), The MIT Press, 2015
- . . . There's a whole web out there