

COURSE: Topics on Machine Learning

INSTRUCTORS: Giorgio Gambosi; Gianluca Rossi
EMAIL: giorgio.gambosi@uniroma2.it
gianluca.rossi@uniroma2.it

COURSE DESCRIPTION

The course provides an introduction to Machine Learning methods and to their application to mining information from datasets. It is strongly related to the courses in Supervised Learning and Unsupervised Learning, with a twofold goal:

1. presenting some additional widely used approaches to classification, clustering or feature selection (such as for example neural networks, support vector machines, non parametric methods, ensemble methods)
2. applying such methods, and some of the ones introduced in those courses, to the analysis of example datasets by coding simple software programs using the Python programming language and some of its packages specifically applied in the framework of Data Science (such as Pandas, Numpy, Scikit-learn, etc.)

An introduction to the programming and to the Python language will be provided at the beginning of the course.

LEARNING OUTCOMES

The course is aimed to provide students the following:

- knowledge of some relevant approaches to data analysis and pattern extraction;
- sufficient expertise in programming, to make them able to implement simple solutions to data science problems in a high level language by applying the methods introduced in this course or in previous ones;
- knowledge of the Python programming language and some of its packages which are most relevant in the framework of data science;
- proficiency in the whole process of dealing with a dataset in a data analysis task, applying different approaches, visualizing their results, comparing them to select the most effective one;

METHODOLOGY

The course is structured in a more theoretical part, devoted to introducing some relevant and widely adopted approaches to mining information in data, and a more practical one, aimed to apply those approaches (and possibly other ones from previous courses) to real datasets.

Time distribution between the two parts will be approximately 50%-50%. The whole course will be laboratory centered, with specific problems proposed to students, organized in groups and suitably monitored/assisted.

ASSESSMENT

Colloquium; weighting: 70%

Project; weighting: 30%

OUTLINE

1. Introduction to programming in Python.
2. Python packages relevant for machine learning.
3. Connectionist approach to machine learning: neural networks.
4. Support vector machines and kernel methods.
5. Ensemble methods.
6. Dimensionality reduction methods.
7. Applying Python in a data science task.

TEXTBOOKS

The main reference for the theoretical part of the course is C.M. Bishop “Pattern Recognition and Machine Learning” Springer, 2007

The main references for the programming/practical one: A.B. Downey “Think Python” O’Reilly, 2012 (freely available at [http:// http://greenteapress.com/wp/think-python/](http://greenteapress.com/wp/think-python/)) for the introduction to Python and S. Raschka “Python Machine Learning” Packt. Publishing Ltd, 2015, for its use in Machine Learning

ADDITIONAL SUGGESTED READING

D.Barber “Bayesian Reasoning and Machine Learning” Cambridge University Press, 2012, freely available at <http://web4.cs.ucl.ac.uk/staff/D.Barber/pmwiki/pmwiki.php?n=Brml.HomePage>

D. McKay “Information Theory, Inference, and Learning Algorithms” Cambridge University Press, 2003, freely available at <http://www.inference.phy.cam.ac.uk/itprnn/book.html>

K.P. Murphy “Machine Learning: a probabilistic perspective” MIT Press, 2012