

Collegio Carlo Alberto

UNIVERSITÀ DEGLI STUDI DI TORINO

MACHINE LEARNING

	Ruggero G. Pensa ruggero.pensa@unito.it
FOTO LECTURER to attach	Roberto Esposito roberto.esposito@unito.it
	Livio Bioglio livio.bioglio@unito.it

Learning Objectives

The aim of this course is to study and apply the most relevant data mining and machine learning models and algorithms to real cases datasets. The attendants will be able to distinguish between descriptive or explorative (unsupervised) and predictive (supervised) methods to extract valuable patterns and models from (large) data.

Important machine learning methods will be introduced theoretically and applied on real datasets using R and Python libraries.

The combination of theory and practice offers students the background necessary to choose the right tool for the job at hand, the expertise to autonomously apply them to real data sets, and the competence to measure the quality of the obtained results.

Course Content

The module focuses on the following method:

- 1) Introduction (data preprocessing, distance metrics, introduction to clustering)
- 2) Clustering Algorithms: k-means, hierarchical clustering, DBSCAN, validation
- 3) Dimensionality Reduction: PCA, SVD, Feature selection
- 4) Factorization-based data clustering: NMF
- 5) Introduction to Neural Networks (NN)
- 6) Model Selection: general principles and applications to NN hyperparameters selection

Course Methodology

Frontal teaching and Lab.

Readings

- Tan, Steinbach, Kumar: Introduction to Data Mining. Addison Wesley
- Ian Goodfellow, Yoshua Bengio and Aaron Courville: Deep Learning

Slides: available during the course

Course Evaluation

Students will be evaluated on individual participation (30%), final exam (70%).

In the final exam they will set up a quick machine learning case study based on some datasets provided by the teachers and will have to answer few questions related to the course program.

About the Instructors

Ruggero G. Pensa received the M.Sc. degree in Computer Engineering from the Polytechnic University of Turin in 2003 and the Ph.D. in Computer Science from INSA of Lyon in 2006. He is Associate Professor at the Department of Computer Science, University of Turin. His main research interests include data mining and knowledge discovery, data science, privacy-preserving algorithms for data management, social network analysis and spatio-temporal data analysis. He is member of the Editorial Board of the Data Mining and Knowledge Discovery journal.

Roberto Esposito received the M. Sc. Degree in Computer Science from the University of Turin in 1999 and then the Ph. D. in Computer Science in 2003. He is Assistant Professor at the Department of Computer Science, University of Turin. His main research interests range in several subfields of Machine Learning including Graphical Models, Ensemble Learning, and Support Vector Machines. He regularly serves as a reviewer for prestigious journals and conferences such as TKDE and IJCAI.

Livio Bioglio received his M.Sc. degree in Computer Science in 2009 at the University of Turin, and his Ph.D. in Computer Science in 2013, in the same institution. In 2013 he was a Post-doc at INSERM, Paris (France), working on complex systems and epidemiology, in particular computational models for analyzing the spread of influenza in realistic populations. Since 2016 he is Post-doc at the University of Turin, studying the diffusion of information in social networks, with focus on privacy concerns.