



UNIVERSITÀ DEGLI STUDI DI PALERMO

DEPARTMENT	Ingegneria		
ACADEMIC YEAR	2024/2025		
MASTER'S DEGREE (MSC)	MANAGEMENT ENGINEERING		
INTEGRATED COURSE	BIG DATA AND ANALYTICS - INTEGRATED COURSE		
CODE	21506		
MODULES	Yes		
NUMBER OF MODULES	2		
SCIENTIFIC SECTOR(S)	ING-INF/05, ING-INF/03		
HEAD PROFESSOR(S)	TINNIRELLO ILENIA	Professore Ordinario	Univ. di PALERMO
OTHER PROFESSOR(S)	LO PRESTI LILIANA	Professore Associato	Univ. di PALERMO
	TINNIRELLO ILENIA	Professore Ordinario	Univ. di PALERMO
CREDITS	6		
PROPAEDEUTICAL SUBJECTS			
MUTUALIZATION			
YEAR	2		
TERM (SEMESTER)	2° semester		
ATTENDANCE	Not mandatory		
EVALUATION	Out of 30		
TEACHER OFFICE HOURS	LO PRESTI LILIANA Tuesday 16:00 17:00 Luogo da concordare via email col docente. Preferibilmente a distanza su piattaforma MS Teams. TINNIRELLO ILENIA Monday 9:00 12:00 Ufficio del docente, presso il DEIM, secondo piano.		

**MODULE
MACHINE LEARNING**

Prof.ssa ILENIA TINNIRELLO

SUGGESTED BIBLIOGRAPHY

- SergiosTheodoridis, Kostantinos Koutroumbas. Pattern Recognition, Academic Press. eBook ISBN: 9780080949123
Hardcover ISBN: 978159749272.
- Paul J. Deitel, Harvey M. Deitel. Introduzione a Python. Per l'informatica e la data science. Pearson 2021
- Dispense del docente.

AMBIT	20929-Attività formative affini o integrative
INDIVIDUAL STUDY (Hrs)	48
COURSE ACTIVITY (Hrs)	27

EDUCATIONAL OBJECTIVES OF THE MODULE

This module presents data-driven techniques for classification problems, based on statistical approaches and models with memory.

More into details, the main objectives of the module are:

- 1) Learn techniques for estimating the probability density or the empirical distribution function of the features used for the classification;
- 2) Apply Bayesian inference in classification problems, including memory-based processes;
- 3) Develop reinforcement-learning techniques.

SYLLABUS

Hrs	Frontal teaching
4	Introduction to the course. Fundamentals of probability and random variables.
8	Classification: hypothesis representation, decision regions, cost functions. Bayesian classifiers. Multi-variate Gaussian distributions. Estimates of probability densities for continuous and discrete features: parametric and non-parametric estimation. Maximum likelihood estimation.
8	Discrete-time Markov processes: transition matrices, steady-state equilibrium conditions, probability limit distributions. Application examples: Google PageRank algorithm. The Viterbi algorithm for classification problems with memory. Markov decision processes and Reinforcement Learning.
Hrs	Practice
7	Introduction to Python and the scikit learn library. Applications of all course concepts to real-world problems of classification and implementation examples.

MODULE DATA ANALYTICS AND STORAGE

Prof.ssa LILIANA LO PRESTI

SUGGESTED BIBLIOGRAPHY

For Python programming:

- Paul J. Deitel, Harvey M. Deitel. Introduzione a Python. Per l'informatica e la data science, Pearson. ISBN: 978-8891915924

Linear and non linear classification, MLP, clustering (mainly Chapters 3, 4, 11 and 13):

- Sergios Theodoridis, Kostantinos Koutroumbas. Pattern Recognition, Academic Press. ISBN: 978-159749272

Pre-processing and data dimensionality reduction:

- Data Mining: The Textbook, 2015, Charu C. Aggarwal, Springer-Verlag New York, ISBN 978-3319141411, disponibile gratuitamente in forma elettronica per gli studenti dell'Ateneo. (freely available online for Unipa students)

Big Data storage:

- Slides fornite dal docente e documentazione ufficiale delle librerie/Slides provided by the teacher and official documentation of the packages

AMBIT	20929-Attività formative affini o integrative
INDIVIDUAL STUDY (Hrs)	48
COURSE ACTIVITY (Hrs)	27

EDUCATIONAL OBJECTIVES OF THE MODULE

This module presents more machine learning techniques and fundamentals of Big Data management systems.

In particular, the main objectives of the module are:

- 1) Learn classification techniques based on linear and nonlinear classifier and on deep learning, and learn clustering techniques.
- 2) Learn the fundamentals of Big Data management systems, programming models for Big Data processing, storage and querying techniques for structured and semi-structured data.
- 3) Implement simple classification, clustering and data analysis technique using the Python programming language.
- 4) Evaluate architectural and algorithmic approaches and determine the more appropriate for the problem at hand.

SYLLABUS

Hrs	Frontal teaching
2	Introduction to the Course. The data analysis process: data collection, pre-processing, application of analysis techniques and knowledge extraction.
3	Data preprocessing: data types, data cleaning, missing data management, sampling.
5	Data dimensionality reduction: Principal Component Analysis, Singular Value Decomposition, Wavelet Transformations, Multi Dimensional Scaling, Graph Embedding.
4	Distances and similarities for different types of data: quantitative data, categorical data, textual data, temporal sequences, graphs.
8	Binary linear classifiers. Perceptron algorithm. Least squares parameter estimation. Linear Support Vector Machine. Training techniques for separable and non-separable classes. Generalizations to multi-class classification problems. Logistic regressor. Linear regressors. Overfitting, underfitting and bias/variance dilemma. Model selection. Non linear classifiers, Decision trees and introduction to random forest.
6	Multilayer neural networks. Dense layers and feature transformations in high dimensional spaces. Activation functions. Loss functions. Training multilayer neural networks: backpropagation algorithm. Introduction to Deep Learning: convolutional layers, pooling operators, dropout, CNN, Autoencoder, Residual neural network, LSTM. Tensorflow library.
2	Clustering: K-Means and hierarchical clustering.
8	Software architectures for Big Data storage: SQL databases, noSQL databases, MongoDB, the MapReduce algorithm, Apache Hadoop, HDFS.
Hrs	Practice
5	Linear and nonlinear classifiers. Deep Learning. Clustering
3	Systems for storage and querying of Big Data.
Hrs	Workshops
8	Review of Python and the numpy, pandas, matplotlib, sklearn packages. Pre-processing and dimensionality reduction. Systems for storage and querying of Big Data.