COMPUTER ENGINEERING (LM55)

(Lecce - Università degli Studi)

Teaching ELEMENTS OF STATISTICAL LEARNING

GenCod A006448

Owner professor ANGELO COLUCCIA

Teaching in italian ELEMENTS OF STATISTICAL LEARNING

Teaching ELEMENTS OF STATISTICAL LEARNING

SSD code ING-INF/03

Reference course COMPUTER ENGINEERING

Course type Laurea Magistrale

Credits 9.0

Teaching hours Front activity hours: 81.0

For enrolled in 2021/2022

Taught in 2022/2023

Course year 2

Language ENGLISH

Curriculum ARTIFICIAL INTELLIGENCE

Location Lecce

Semester First Semester

Exam type Oral

Assessment Final grade

Course timetable https://easyroom.unisalento.it/Orario

BRIEF COURSE DESCRIPTION	The course provides a broad coverage of the essential elements of statistical learning as well as concepts, methodologies and tools that find application in machine learning, data science, and related data-driven fields.
REQUIREMENTS	Basics of Probability and Statistics, Mathematics

COURSE AIMS

Knowledge and understanding. Students must have a solid background of statistical techniques, including probability and stochastic processes, that can be applied to solve problems in engineering with a data-driven approach. They should be able to:

- Describe the characteristics of advanced statistical learning techniques and discuss the principles of data science and machine learning design;
- Understand the different types of techniques that can be exploited to solve regression, classification, and other learning problems;
- Describe how traditional machine learning algorithms and (deep) neural networks can be suited to different types of problems.

Applying knowledge and understanding. After the course the student should be able to:

- Work with analytical models and solve optimization, classification, and estimation problems related to the course topics;
- Describe the peculiar aspects and main challenges of machine learning, and how advanced statistical techniques can be adopted to efficiently cope with them;
- Understand the differences among several techniques addressing the same problem and recognize the main trade-offs;
- Discuss the evolution of the data-driven paradigm, the related ongoing trends and risks.

Making judgements. Students are guided to learn critically what is taught during classes, comparing different approaches, while having a clear view of the big picture.

Communication. It is essential that students are able to communicate with a varied and composite audience, not culturally homogeneous, in a clear, logical and effective way, using the methodological tools acquired, their scientific knowledge, and the specialty vocabulary. The course promotes the development of the following skills: ability to highlight and expose in precise terms the characteristics of a variety of statistical and machine learning concepts and techniques; ability to describe and analyze the different options available for a given application scenario or use case, and illustrate the main trade-offs; ability to communicate in a rigorous way backed by statistical reasoning and data science knowledge.

Learning skills. Students must acquire the critical ability to discuss, with originality and autonomy, the most important aspects of statistical (machine) learning and, in general, cultural issues linked to data science especially in the ICT domain. They should be able to develop and apply the knowledge learned in the continuation of their studies and in the broader perspective of cultural and professional self-improvement of lifelong learning. Therefore, students are asked to refer to and compare different sources and textbooks, possibly by also autonomously selecting authoritative materials from the vast amount of information available (libraries, online repositories, and the Web at large).

TEACHING METHODOLOGY	**Teaching Methods.** The course aims at enabling students to understand statistical learning theory and data-driven methods, keeping an unified view and being able to navigate the complexity of modern scenarios. This will be done using the following teaching method. Every concept or technique will be introduced in terms of motivations, technical peculiarities, and application scope. The presentation of each topic will be linked to the background studied in previous courses, and continuously connected to the preceding and subsequent topics within the present course. The course consists of frontal lessons with slides and blackboard, together with class exercises. There will be theoretical lessons, qualitative discussions, and examples about how knowledge is put into practice in real cases. A part of the lessons will be also devoted to illustrate related ongoing research directions in the field.
ASSESSMENT TYPE	Written and/or oral. The final (typically written) exam consists of questions aimed at verifying to what extent the student 1) has gained knowledge and understanding of the selected topics of the course, 2) is able to discuss complex aspects in a synthetic way, and 3) has gained adequate degree of maturity in linking concepts within a system view. Small exercises may be included in the questions so that the student can demonstrate his/her ability to 1) correctly adopt formal techniques for solving well-defined problems, and 2) integrate different concepts and tools.
OTHER USEFUL INFORMATION	**Office Hours** By appointment; contact the instructor by email or at the end of class meetings.

FULL SYLLABUS

Introduction to Machine Learning, recapitulation of Probability and Stochastic Processes

Learning theory for parametric models
(linear regression,statistical decision theory and classification, bias, MSE, trade-off, model complexity, Maximum Likelihood, Bayesian inference, curse of dimensionality, cross-validation, MSE linear estimation and applications, stochastic gradient descent, least-squares approach)

Overview of Supervised Learning
(Least Squares and Nearest Neighbors,local methods in high dimensions, statistical models, supervised learning and function approximation, model selection and the bias-variance trade-off)

Classification methods
(Bayesian classification, the Nearest Neighbor rule, logistic regression, Fisher's linear discriminant, classification trees and bagging, the boosting approach, random forests)

Sparse signal representation and learning
(LASSO, compressed sensing, embeddings, ensemble learning)

Learning in Reproducing Kernel Hilbert Spaces
(Kernel smoothers and regression, representer theorem, kernel ridge regression, support vector machines)

Unsupervised learning
(clustering, principal components and dimensionality reduction)

Bayesian learning
(regression: a Bayesian perspective, Occam's razor rule, the exponential family and the Maximum Entropy principle, latent variables and the EM algorithm, Gaussian mixture models and clustering, Gaussian processes)

Monte Carlo methods
(random number generation and sampling, Monte Carlo methods and the EM algorithm, Markov chain Monte Carlo methods and the Metropolis method, Gibbs sampling, hidden Markov models, particle filtering and Kalman filtering)

Neural Networks and Deep Learning
(perceptron, the backpropagation algorithm, universal approximation, neural network architectures, deep autoencoder)

Ongoing trends and risks in data-driven approaches

REFERENCE TEXT BOOKS

Textbooks (other specific references are provided during the course)

S. Theodoridis, "Machine Learning: A Bayesian and Optimization Perspective", 2nd edition, Academic Press, 2020

T. Hastie, R. Tibshirani, J. Friedman, "The Elements of Statistical Learning: Data Mining, Inference, and Prediction", 2nd edition, Springer, 2009

O. Simeone, "A Brief Introduction to Machine Learning for Engineers", Foundations and Trends in Signal processing, Now publishing, 2021