# Stochastic Processes and Optimization for Machine Learning

## Prof. Francesco Ballarin; Prof. Bruno Buonaguidi

***COURSE AIMS AND INTENDED LEARNING OUTCOMES***

This course covers two broad topics which are at the basis of modern machine learning techniques, namely stochastic processes and optimization methods. The aim is to develop knowledge on the fundamental aspects of such topics, as well as a practical understanding of how they work and why they are used in combination with machine learning techniques. Since many modern methods in Data Analytics are built on the aspects developed in this course, the ultimate goal is to provide students with the tools they need to understand the core concepts of stochastic processes and optimization methods, and further develop them in subsequent courses as well as in their professional career.

By the end of the course, students are expected to:

1. have acquired solid knowledge of the methodological aspects, as well as practical realization, of the theory at the basis of stochastic processes and optimization methods (*Knowledge and understanding*);
2. be able to apply the main results of the stochastic processes theory in the solution of real-life problems and exercises. Furthermore, they are expected to be able to implement in Python a broad range of algorithms related to optimization methods and recognize how such algorithms are used in modern machine learning techniques. Students will be able to apply these quantitative tools to problems arising in Economics and Business (A*pplying knowledge and understanding*);
3. be able to model phenomena of the social, management and computer sciences by means of appropriate stochastic models. Furthermore, they are expected to be able assess the performance of the studied optimization algorithms, and to be able to choose among different optimization method depending on the characteristics of the problem at hand (M*aking judgements*);
4. be able to describe with an appropriate language the mathematical description of methods in stochastic processes and optimization, the underlying assumptions behind them, and how and why such methods are used in modern machine learning techniques (C*ommunication skills*);
5. have acquired the main concepts and tools to independently learn and develop statistical and computational methods for Business Analytics (*Learning skills*).

***COURSE CONTENT***

Module A - Optimization*(Prof. Francesco Ballarin)*

* Introduction: optimization formulation of data science problems, optimality conditions, classes of functions.
* First order methods: gradient descent method, line search techniques, convergence rates, acceleration techniques.
* Second order methods: Newton method, Quasi-Newton methods, diagonal scaling.
* Stochastic methods: stochastic gradient method, mini-batch stochastic gradient method.
* Regularization techniques: ridge & LASSO regularization, subgradient method, proximal gradient method.
* Introduction to feedforward neural networks, and training of a neural network with the help of automatic differentiation techniques.
* Lectures for each of the previous items will introduce the numerical methods, as well as implement them in Python in order to discuss applications motivated by data science use cases, including linear regression, logistic regression, neural networks.

Module B - Stochastic processes *(Prof. Bruno Buonaguidi)*

* Markov Chains.
* The Exponential Distribution and the Poisson Process.
* Continuous-Time Markov Chains.
* An introduction to Brownian motion.
* Examples of Gaussian processes in machine learning.

***READING LIST***

Class notes, slides, coding and further material will be posted on the University platform Blackboard.

 Further suggested readings are:

* S. M. Ross (2014). Introduction to probability models, 11th edition, Elsevier (Chapters 4, 5, 6, 10)
* C. E. Rasmussen and C. K. I. Williams (2006). Gaussian Processes for Machine Learning, MIT Press.
* J. E. Gentle (2009). Computational statistics, Springer (Chapter 6).
* M. P. Deisenroth, A. A. Faisal, C. S. Ong (2020). Mathematics for Machine Learning, Cambridge University Press.
* C. Aggarwal, “Linear Algebra and Optimization for Machine Learning”, Springer, 2020.
* I. Goodfellow, Y. Bengio, A. Courville, “Deep Learning”, MIT Press, 2016.

***TEACHING METHOD***

A blend of lectures (15 hours in Module A, 30 hours in Module B) and exercise sessions with Python (15 hours in Module A). Attending the lectures and exercise sessions, active participation and ongoing personal study are strongly recommended.

***ASSESSMENT METHOD AND CRITERIA***

The assessment of Module A is based on two parts:

1. Homework (max 14 marks): computational homework will be assigned after exercise sessions in Python, and will be carried out individually by the student;
2. Written examination (max 18 marks) with open-ended questions on optimization methods, their application in data science, and exercises.

The assessment of Module B is based on a written examination with open-ended questions on methods and exercises.

The final grade is the average of the grades achieved in the two modules.

In addition to the usual exam sessions, students will be offered the opportunity to take the written examination of Module A after the end of the lectures of Module A.

***NOTES AND PREREQUISITES***

Students enrolling in this course are expected:

* for Module A, to have basic knowledge in computational statistics at the level of the corresponding course taught during the first year and a fair knowledge of the Python language
* for Module B, to have knowledge of probability at the level of the “Mathematical Methods and Probability” course taught in the first year.