Statistical Learning Theory and Applications
Prof. Lorenzo Rosasco e Prof. Silvia Villa

Period: 30/01, 31/01, 01/02
Duration: 12 h (6 lessons)

Course description

The course covers foundations and recent advances of machine learning from the point of view of statistical learning and regularization theory.

Understanding intelligence and how to replicate it in machines is arguably one of the greatest problems in science. Learning, its principles and computational implementations, is at the very core of intelligence. During the last decade, for the first time, we have been able to develop artificial intelligence systems that can solve complex tasks, until recently the exclusive domain of biological organisms, such as computer vision, speech recognition or natural language understanding: cameras recognize faces, smart phones understand voice commands, smart speakers/assistants answer questions and cars can see and avoid obstacles. The machine learning algorithms that are at the roots of these success stories are trained with examples rather than programmed to solve a task.

Among different approaches in modern machine learning, the course focuses on a regularization perspective and includes both shallow and deep networks. The content is roughly divided into two parts. In the first part, key algorithmic ideas are introduced, with an emphasis on the interplay between modeling and optimization aspects. Algorithms that will be discussed include classical regularization networks (regularized least squares, SVM, logistic regression), stochastic gradient methods, implicit regularization, sketching, sparsity based methods and deep neural networks. In the second part, key ideas in statistical learning theory will be developed to analyze the properties of the various algorithms previously introduced. Classical concepts like generalization, uniform convergence and Rademacher complexities will be developed, together with topics such as bounds based on margin, stability, and privacy. The final part of the course focuses on deep learning networks. It will introduce an emerging theoretical framework addressing three key puzzles in deep learning: approximation theory -- which functions can be represented more efficiently by deep networks than shallow networks -- optimization theory -- why can stochastic gradient descent easily find global minima -- and machine learning -- whether classical learning theory can explain generalization in deep networks. It will also discuss connections with the architecture of visual cortex, which was the original inspiration of the layered local connectivity of modern networks and may provide ideas for future developments of deep learning.

The goal of the course is to provide students with the theoretical knowledge and the basic intuitions needed to use and develop effective machine learning solutions to challenging problems.

Prerequisites
We will make extensive use of basic notions of calculus, linear algebra and probability. We will introduce a few concepts in functional/convex analysis and optimization. Students are also expected to have basic familiarity with MATLAB/Octave.