CS-605

Computational and statistical learning theory

Srebro Bartom Nathan

Cursus	Sem.	Type
Computer and Communication Sciences		Opt.

Language of teaching	English
Credits	3
Session	
Exam	Written
Workload	90h
Hours	**56**
Courses	28
Exercises	28
Number of positions	**25**

Frequency

Only this year

Summary

Statistical learning theory for supervised learning and generalization in PAC and online models (VC theory, MDL/SRM, covering numbers, Radamacher Averages, boosting, compression, stability and connection with strong convexity in Banch spaces); Computational tractability of learning.

Content

The purpose of this course is to gain a deeper understanding of machine learning by formalizing learning mathematically, studying both statistical and computational aspects of learning, and understanding how these two aspects are inseparable. The course is intended both for students interested in using machine learning methods and that would like to understand such methods better so as to use them more effectively, as well as for students interested in the mathematical aspects of learning or that intend on rigorously studying or developing learning algorithms.

We will discuss classic results and recent advances in statistical learning theory, touch on computational learning theory, and explore the relationship with stochastic optimization and online regret analysis. Our emphasis will be on concept development and on obtaining a rigorous quantitative understanding of machine learning. We will also study techniques for analyzing and proving performance guarantees for learning methods.
Note: This is not intended as an introductory course on machine learning, and is recommended mostly for students familiar with machine learning problems and methods, but might be appropriate also for mathematically inclined students wishing a formal view of machine learning.Specific Topics
The Statistical Model (Learning Based on an IID Samples):
The PAC (Probably Approximately Correct) and Agnostic PAC models.
Generalized Learning and relationship to Stochastic Optimization
VC Theory: Cardinality, VC dimension, a the growth function
Description Length Bounds, Structural Risk Minimization
Uniform Learning and No-Free Lunch Theorems
PAC-Bayes
Compression Bounds
Scale sensitive classes: Fat Shattering Dimension, Covering Numbers and Radamacher Averages
Tight Characterization of Learning in terms of the VC and Fat Shattering Dimensions
Relative bounds, "Optimistic" rates, and Local Rademacher Analysis
Boosting, including three different views of AdaBoost (traditional PAC view; learning a linear predictor with coordinate descent; boosting the margin and l1 as its dual)

Online Learning, Online Optimization and Online Regret:
The Perceptron Rule and Online Gradient Descent
Experts and the Winnow Rule
Strong convexity, stability, and online learning in Banach spaces as a unifying theory.
Bregman Divergence and Online Mirror Descent
Online to Batch Conversion

Computational Lower Bounds:
Computational Hardness of Proper Learning
Cryptographic Hardness of Learning
Deep Learning through the lens of Learning Theory

Note

Learning outcomes:
By the end of the course, the student must be able to:
Formally understand fundamental concepts in machine learning, understand learning guarantees and impossibility results, use standard analysis techniques

Learning Prerequisites

Required courses

probability, algorithms, introduction to machine learning