Massachusetts Institute of Technology
Department of Electrical Engineering and Computer Science

6.435: Theory of Learning and System Identification

Description

Given a finite set of Data:

$$\{x_1, x_2, \ldots x_n, y_1, y_2, \ldots y_n\},$$

how can we derive the functional dependence in this data in terms of a "dynamic" model that is both parsimonious and predictive?

The answer to this question is quite elusive unless we add more assumptions on "how" the data is generated. Many disciplines address various sets of assumptions on the data. If the pairs (x_i, y_i) are assumed probabilistic and IID, then statistical learning theory addresses the above question through the notion of Risk Minimization, which can be viewed as a generalization of the problem of density estimation. Within this theory, the functional dependence is hypothesized through the selection of a class of functions, whose *dimension* in addition to a notion of Empirical Risk Minimization, control the predictive power of the derived model. If the data set is dependent, then time-series analysis using linear models offers a powerful paradigm for deriving a predictive model from data. The theory of Hidden Markov Models also plays an important role under this assumption. In all such cases, there is always the question of model selection. Notions related to "Description Length" and "Kolmogorov Complexity" play a central role in answering such a question.

In this course, we present a unified framework that addresses the above problem. We demonstrate that the notion of Empirical-Risk Minimization encompasses a variety of standard statistical inference problems including *Function Approximation*, *Regression Analysis*, *Density Estimation*, and *Prediction-Error Minimization*. We show that the notion of *Uniform Convergence* of Empirical Risk is central to all such problems. We derive conditions for such convergence in terms of the *Complexity* of the model set. In particular, we derive:

- ▹ finite-data bounds on the difference between Empirical Risk and actual Risk in terms of the data length and the complexity of the model set known as *VC dimension*,
- ▹ asymptotic bounds and asymptotic distribution of estimated parameters in the case of linear models and Hidden Markov Models for which we bring out the role of *Relative Entropy* and the notion of input selection and persistence of excitation,
- ▹ a complexity notion that can be used to evaluate model sets known as *Description Length* and its relationship to *Structured Risk Minimization*,
- ▹ algorithms for constructing such models including Support Vector Machines, Least Squares, Maximum Likelihood, and Expectation Minimization, and
- ▹ tools for experiment design.

While the course presents a rigorous treatment of this subject, most proofs will be done for *Indicator Functions* in the case of independent data (which emerge in applications in Pattern Recognition), and Guass-Markov Models and Hidden Markov Models in the case of dependent data. Hidden Markov Fields and Graphical Models will also be addressed.

Course Details

General Information

Instructors: Prof. Munther A. Dahleh (`dahleh@mit.edu`), Room 32-D734, x3-3892 and Prof. Sanjoy K. Mitter (`mitter@mit.edu`), Room 32-D562, x3-2160.

TA: Mesrob I. Ohannessian (`mesrob@mit.edu`), Room 32-D740, x4-1544.

Secretary: Ms. Fifa Monserrate (`fifa@mit.edu`), Room 32-D733, x3-2184.

Lectures: TR 2:30-4pm, Room 56-114.

Prerequisites: 6.437 or 6.438, or 6.867 or approval of instructors.

Grading: Five homework sets and scribing duty, equally weighed (75%). Final project (25%).

Homework Policy

The homeworks are *not* intended as tests, but as vehicles for learning, complemented by the homework solutions that we hand out, and by any discussions that you have about the problems. *Moderate* collaboration on homework with your *classmates* is permitted. Discussions with the teaching staff are encouraged. There is no harm in seeking minor assistance from others who are knowledgeable but not involved in the class, although we would much prefer that your discussions be with those in the class.

We expect each of you to put in enough time *alone* to understand the specific difficulties and issues raised by each homework problem. We also expect that you will *independently* write up the actual solutions that you turn in, and not give us direct copies of a classmate's solutions! You should note on your solutions the names of those you have collaborated with or obtained help from.

Email Announcements and Website

It is important that you give us an e-mail address for yourself, and that you also check it regularly, as there will quite likely be administrative and other announcements sent out from time to time by the 6.435 teaching staff. Course material will be posted on the usual web site, at: http://web.mit.edu/6.435/www/.

Recommended Texts

1. Vapnik, V. N., *Statistical Learning Theory*, Wiley-Interscience, 1998.

2. Vidyasagar, M., *Learning and Generalization: With Applications to Neural Networks*, Springer, e1: 1997, e2: 2003.

3. Ljung, L., *System Identification: Theory for the User*, Prentice Hall, e1: 1987, e2: 1999.

4. Jordan, M. and Bishop, C., *Introduction to Graphical Models*, (unpublished, will be made locally available.)

5. Cover, T. M. and Thomas, J. A., *Elements of Information Theory*, Wiley-Interscience, e1: 1991, e2: 2006.

Tentative Schedule

The following is a tentative schedule of the lecture topics, with respective references.

Date, Day		Topic	Reference
Feb 6, T	L1:	Introduction to Learning and System Identification	Notes
8, Th	L2:	Risk Minimization. Examples: Indicator functions, Regression, Density estimation. Implications. Support Vector Methods. **PS1 handed out.**	Vapnik, Notes
13, T	L3:	Entropy, Differential Entropy, and Relative Entropy. KL distance. Large Deviation and Chernoff Bounds. Implications to PMF estimation of Discrete Random Variables	Cover
15, Th	L4:	Density Estimation and Glivenko-Cantelli Theorem. Relations to Risk Minimization. Uniform Convergence of Empirical Risk. PAC Learning. Notion of Entropy of a model set.	Vapnik, Vidyasagar
20, T		Monday Schedule	
22, Th	L5:	Bounds on Uniform Convergence. Distribution-Free Bounds and VC-Dimension. Examples of model sets with finite VC-dimension. **PS1 turned in, PS2 handed out.**	Vapnik, Vidyasagar
27, T	L6:	Growth bounds in terms of VC. Implications to Density Estimation.	Vapnik
March 1, Th	L7:	Extension of results to general functions (no proofs).	Vapnik, Vidyasagar
6, T	L8	Model Evaluation: Coding theory, Kraft's inequality, Typical sets. Introduction to MDL	Cover
8, Th	L9:	MDL and Structured Risk **PS2 turned in, PS3 handed out.**	Notes
13, T	L10:	Algorithms, SVMs, Stochastic Algorithms and convergence	Notes
15, Th	L11:	Dependent samples: Dynamic Models. Gauss-Markov Models. HMMs.	Notes
20, T	L12	ARX, Prediction Error and relations to Risk Minimization. Convergence. Asymptotic Distributions. persistence of Excitations.	Ljung
22, Th	L13:	General Models. Uniform Convergence results (TBD). **PS3 turned in, PS4 handed out.**	
27, T - 29, Th		Spring Vacation	

Date, Day		Topic	Reference
April 3, T	L14:	Discrete Stochastic Models: HMM. Probability function. Prediction, Smoothing. Backward and Forward Algorithms. Viterbi and DP.	
5, Th	L15:	ML Estimation. Expectation Minimization. Max-Sum, Max-Product algorithms. Baum-Welch algorithm. Segmental K-means. **PS4 turned in, PS5 handed out. Project Proposals Due.**	Notes
10, T	L16:	Convergence of Baum-Welch and Segmental K-means.	Notes
12,Th	L17:	Markov Fields. Graphical models. Density decomposition. Bayesian Networks.	Jordan
17, T		Holiday	
19, Th	L18:	Belief propagation. Message Passing. EM algorithm. **PS5 turned in.**	Jordan
24, T	L19:	Gibbs Distributions. Hammersly-Clifford equivalence between Markov Fields and Gibbs Distributions.	Jordan
26, Th	L20:	Exponential Families. Geometry. Graphical models. (Drop Date)	Wainwright
May 1, T	L21:	Exponential Families: Variational Methods. Duality	Wainwright
3, Th	L22:	More on Duality	Wainwright
8, T	L23:	Sample-path view of learning.	Kulkarni
10, Th	L24:	Sample-path view of learning.	Kulkarni
15, T	-	Projects	
17, Th	-	Projects	