Statistical Learning Theory 895231

Uri Shaham

Academic Year	
Semester	A
Year of Study	
Course Frequency (If relevant)	Annual
Credit Hours	3
Type of course	elective

Course description/summary (Short abstract)

Statistical Learning Theory studies the foundations of machine learning: **why and when learning from data works**, how many samples are needed, and how model choices affect generalization beyond the training set. The course develops a rigorous framework for supervised learning (classification and regression), focusing on **generalization bounds**, **sample complexity**, and **capacity measures** such as VC dimension and data-dependent complexities (e.g., Rademacher complexity). Students learn to reason formally about empirical risk minimization, overfitting, regularization, margins, and kernel methods, and to connect theoretical results to practical design decisions (model selection, validation strategies, and robustness). By the end of the course, students will be able to derive and interpret core learning guarantees, compute or bound capacities for common hypothesis classes, and critically evaluate claims about model performance using principled statistical arguments. The course also offers a brief view of modern questions in learning theory (e.g., overparameterization and implicit bias), helping students read research-oriented material and build a foundation for advanced machine learning, optimization, and theoretical data science.

Prerequisites

- Probability
- Linear Algebra 2
- Machine Learning

Course Objectives & Learning Outcomes

By the end of the course, students will be able to:

- Formulate supervised learning problems using a precise statistical framework (hypotheses, losses, risks).
- Explain and apply the core mechanisms that enable generalization: concentration, uniform convergence, stability, and complexity control.
- Use learning-theoretic tools to justify modeling choices (hypothesis class selection, regularization, margins, kernels).
- Communicate theoretical results clearly and relate them to empirical behavior through small experiments and written reasoning.

Upon successful completion, students can:

1. **Define** learning models (ERM, PAC-style guarantees), risks, losses, and hypothesis classes.
2. **Apply** concentration inequalities and union bounds to obtain finite-sample statements.
3. **Compute/analyze** VC dimension (or meaningful bounds) for standard classes (intervals, halfspaces in low dimensions, linear separators with margin assumptions, etc.).
4. **Derive and interpret** generalization bounds (VC-based and data-dependent, as introduced).
5. **Compare** regularization viewpoints: complexity control, stability, and optimization-based interpretations.
6. **Explain** kernel methods at a theoretical level (feature maps, RKHS intuition, representer-type reasoning) and connect to margin-based generalization.
7. **Design** a small empirical study (simulation) that illustrates a theoretical concept (e.g., overfitting vs. capacity, margin effects).
8. **Write** concise solutions and proofs for core results at an undergraduate-math level.

Weekly Schedule

Lesson #	Content	Assignments	Reading
1	Course overview; supervised learning setup; loss, risk, empirical risk; overfitting & generalization, Hoeffding-bound, union bound		Course notes: Learning setup + examples UML Ch. 1
2	**PAC-style learning** perspective; sample complexity intuition; realizable vs. agnostic for finite hypothesis classes		UML Ch. 2-3
3	Uniform convergence & ERM; VC dimension: shattering	HW1 released	UML Ch. 4,6
4	Growth function; Sauer lemma; The fundamental theorem of learning theory		UML Ch. 6
5	**The Bias-Complexity Tradeoff:** No-Free-Lunch Theorem. Error decomposition (Approximation vs. Estimation error).		UML Ch. 5
6	**Non-Uniform Learnability:** Structural Risk Minimization (SRM). Minimum Description Length (MDL).	HW2 release	UML Ch. 7
7	Data-dependent complexity: Rademacher complexity intuition and basic bounds		Notes
8	Covering numbers		Notes

9	**Boosting:** Weak vs. Strong learning. AdaBoost algorithm and its theoretical bounds.	HW3 release	UML Ch. 10
10	**Universal approximation of Neural nets:** Cybenko's theorem, Barron's theorem.		Notes
11	**Generalized Linear Models**		Notes
12	**Regularization & Stability:** Tikhonov regularization. Stability-based generalization bounds. **Online learning viewpoint**: regret	HW 4 release	UML Ch. 13
13	**Modern theory: Generalization in the overparametrized regime**		Notes

Assessment Methods and Grading Composition

- **Problem Sets (4 total)** – 40%
- **Exam** – 60%

Bibliography

Mandatory Reading:

[UML] Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press. (This is the primary text).

Additional Reading (Elective):

- **[ESL]** Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.

- **[FML]** Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of Machine Learning*. MIT Press.