# Lecture schedule

Updated April 22, 2019

The tentative lecture schedule for the semester is to discuss material from the course notes. The chapters will be followed in order with the following exceptions: material from Chapter 3 (minima, convexity, strong convexity, smoothness) and Chapter 4 (functions determined by kernels) will be discussed as needed for later chapters. Only selections from Chapters 10-12 will be discussed.

Tu 1/15 Course mechanics, main goals of learning, examine the programming problem (last problem) for problem set 1. (It is about the Iris flower data set and performance of the K Nearest Neighbor classifier. It demonstrates role of training data and test data for supervised learning.) Quick run through Chapter 1, focusing on a sequence of independent tosses of a biased coin. Three distinct problems discussed: estimation, prediction, and learning.

Th 1/19 Sections 5.1-5.2. Concept learning in the realizable case, PAC learnability. PAC learnability for finite hypothesis classes and for the axis parallel rectangles example (realizable case).

F 1/20 (optional recitation session) – discussion of Theorem 5.1 in the course notes – realizable concept learning with finitely many hypotheses are PAC learnable. Also, Section 3.3 (convex functions) and Appendix A (especially Jensen's inequality).

Tu 1/22 Chapter 2. Concentration inequalities

Th 1/24 Finish up concentration inequalities. A converse result (aka no free lunch theorem), realizable case (This is a problem in Problem set 1. A variation may be found in Chapter 5 of *Understanding Machine Learning*.)

F 1/25 (optional recitation session) – further discussion of homework problem on converse result. examples using concentration inequalities

Tu 1/29 (ps 1 due) Section 5.3-5.5 (also see Section 5.1.2) Formulation of the learning problem, continued. Topics include agnostic (also called model free) learning of functions or concepts and the extension of the PAC framework for such problems. As an example, a model free concept learning problem is defined by starting with a realizable one and then assuming the labels are flipped with probability $\eta$. The mismatched minimization lemma was briefly discussed. empirical risk minimization (ERM) and the related mismatch minimization lemma.

Th 1/31 Complete Chapter 5, begin Chapter 6: Empirical risk minimization (ERM) and observation that UCEM (uniform approximation of empirical moments) implies PAC property. Chapter 6 is about Empirical Risk Minimization and Rademacher averages. Section 6.1 gives a cleaner more general formulation of the learning problem geared towards study of ERM. Section 6.2 shows the mean of maximum deviation of empirical averages can be bounded above by mean Rademacher averages. Combining such bound on the mean with the concentration inequality technique of Chapter 2 provides a path to proving that ERM is PAC if Rademacher averages can be suitably bounded.

F 2/1 (optional recitation session) Review of problems from problem set 1.

Tu 2/5 Chapter 6: Sections 6.3, 6.4. The first part of Chapter 6 naturally leads us to look more closely at Rademacher averages (Section 6.3 to start with, and then into Chapter 7 on VC dimension.) Section 6.4 gives a sneak preview of how VC dimension (next chapter) combines with Rademacher averages.

Th 2/7 and F 2/8 – No lecture or recitation due to CSL Student Conference https://studentconference.csl.illinois.edu

Tu 2/12 Chapter 7: VC dimension, examples including Dudley class, ending with bound on the Rademacher average of a class of binary valued functions in terms of the VC dimension of the class (based on the finite class lemma for Rademacher averages and the Sauer-Shelah lemma).

Th 2/14 Section 8.1 Binary classification: putting together the previous theory to derive the fundamental theorem of concept learning. (Also, discussed proof of Sauer-Shelah lemma based on shifting algorithm, upper bound on VC dimension for Dudley class, and discussed problem set 2 problems.)

F 2/15 (optional recitation session) Review of Rademacher averages and VC dimension

Tu 2/19 (ps 2 due) Section 8.1 - 8.2 Review of Section 8.1, fundamental theorem of concept learning, pointing out Lemma 8.1 ($V(\mathcal{C}) = V(\mathcal{F}_{\mathcal{C}})$). Section 8.2 on use of surrogate loss function, and discussion of contraction principle for Rademacher averages, Section 6.3.

Th 2/21 Sections 8.3 and 8.4 Weighted linear combinations of classifiers, and the ada boost algorithm. Uses some properties of Rademacher averages from Section 6.

F 2/22/19 (optional recitation session) Review/introduction to Hilbert spaces (Section 4.1)

Tu 2/26 Section 8.5, Sections 4.2-4.3. Bounding Rademacher averages for neural networks, introduction to the use of kernels and reproducing kernel Hilbert spaces (RKHS)

Th 2/28 Sections 4.2-4.3, Section 8.6, kernels as measures of complexity of functions, representer theorem (connected with the use of data dependent bases and the kernel trick in machine learning)

F 3/1 (optional recitation session) Overview of problem set 3 and/or review for exam 1.

Tu 3/5 (ps 3 due) Section 8.7, Convex risk minimization, Chapter 9: Regression with quadratic loss

W 3/4 EXAM 1, 7-9 pm, 2013 ECEB

Th 3/7 Chapter 9: Regression with quadratic loss, continued, Sections 3.4 & 3.5 Strongly convex and smooth convex functions

F 3/8 (optional recitation session) Zeyu discussed a convergence result for gradient descent assuming smooth convex function

Tu 3/12 Chapter 10: Sections 10.1-10.2 Consistency from generalizability plus asymptotic ERM, stability for replace one perturbations of data under strong convexity

Th 3/14 Sections 10.3 - 10.4 Equivalences of generalizes and stable; stability of stochastic gradient descent

F 3/15 (optional recitation session) Bolton reviewed problems on problem set 4

SPRING BREAK

Tu 3/26 (ps 4 due) Section 10.5 Analysis of stochastic gradient descent (SGD)

Th 3/28 Complete Section 10.5, begin Chapter 11, Section 11.1 Online convex programming and a regret bound (Zinkevich)

F 3/29 (optional recitation session) Zeyu discussed following from 2017: Problem 1 of ps6, Problem 2 of exam 2

Tu 4/2 Sections 10.1-10.2 Complete proof of Zinkevich result, apply it to get $B^2L^2$ bound for classical perceptron algorithm

Th 4/4 Section 10.3 On the generalization ability of online learning algorithms

F 4/5 (optional recitation session) TBD

Tu 4/9 (ps 5 due) Chapter 11 Minimax lower bounds – an information theoretic approach (next four lectures)

Th 4/11 Chapter 11 Minimax lower bounds (continued)

F 4/12 (optional recitation session) On the use of Fano's inequality to prove the converse of Shannon's theorem in information theory

Tu 4/16 Chapter 11 Minimax lower bounds (continued)

Th 4/18 Chapter 11 Minimax lower bounds (continued)

F 4/19 (optional recitation session) Discussion of problem set 6

Tu 4/23 Chapter 13, Empirical vector quantization

Th 4/25 (ps 6 due) Chapter 14, Dimensionality reduction in Hilbert spaces

F 4/26 (optional recitation session) Review for Exam 2 using old exams?

Tu 4/30 Exam 2, 7-9 pm, 2013 ECEB

Presentations to be scheduled Monday-Wednesday, May 6-8, papers due Friday, May 10, 5pm.