APPM 7400 Theory of Machine Learning Spring 2020

Websites: canvas and github.com/stephenbeckr/ML-theory-class

Instructor and contacts: Stephen Becker, ECOT 338, (303) 492-0662, *stephen.becker@colorado.edu*. I will respond to most reasonable email, but (1) lengthy questions should be posed at office hours and not via email, and (2) I will not respond to email at odd hours or just before homework is due.

Meeting times: Monday, Wednesday and Friday, "Newton lab" (ECCR 257), 1:00 PM to 1:50 PM.

Office hours: Wednesdays 4–5 PM, Thursdays 12:45 PM – 2:45 PM, in ECOT 338.

Course Format: This is a lecture-based course, and the instructor will present proofs. The textbooks provide supplementary material and homework problems. Students are expected to read the book as necessary to fill in gaps not covered in lecture. Reading the book before lecture is useful but not required. Homeworks will require students to write proofs, synthesizing concepts learned from lecture and the book.

Estimated Workload: This is a 3-credit-hour class. It is more advanced than 5000-level graduate classes, but intended to have a lower work-load than core 5000-level classes (such as classes used to prepare for PhD preliminary exams). The actual work-load will of course depend on the student's background. Note that there are no projects, so workload is roughly uniform throughout the semester.

Prereqs: Undergraduate analysis (APPM 4440 "Real Analysis" or equivalent, or even better, graduate analysis like APPM 5440) and mathematical maturity (and we assume the usual prereqs for analysis, such as linear algebra and probability), as this is a proofs-based math class. Familiarity with machine learning algorithms is very helpful (e.g., CSCI 5622).

Text: The main textbook is Understanding Machine Learning: From Theory to Algorithms, 1st edition, by Shai Shalev-Shwartz and Shai Ben-David (Cambridge University Press, 2014, ISBN-13: 978-1107057135, $35–$45 on Amazon).

As a supplemental text, we will use Foundations of Machine Learning, 2nd edition, by Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar (MIT Press, 2018, ISBN-13: 978-0262039406, $50 on Amazon). The authors host a free PDF of the book at their website.

Syllabus: See next pages for detailed syllabus

Recitations: There are no recitation session.

Exams: There will be an in-class midterm exam (10% of overall class grade; Friday, March 6) and a final exam in our allotted final exam time (10% of overall class grade). The final exam is Monday, May 4 2020 4:30 p.m.–7 p.m.

Project: No projects

Grades/Homeworks: Your grade will be mostly based on homeworks (80%). These will usually be due every two weeks, but sometimes every week (in which case they will be shorter).

Late homework and cheating policy: Homework is due at the beginning of class. Late homework will not be accepted or graded. Any instance of cheating or other violation of the honor code will result in a 0 grade on the relevant assignment and a referral to the honor council.

Course web page: It is your responsibility to check the Canvas page on a regular basis, as well as the github website.

Dropping the course: Advice from your department advisor is recommended before dropping any course. After Jan. 30, dropping a course results in a "W" on your transcript and you'll be billed for tuition. After March 22, dropping the course is possible only with a petition approved by the Dean's office.

Syllabus

A proofs based course on the underlying theory behind machine learning. Instead of chasing the latest-greatest algorithm, the course asks fundamental questions about what learning means and what can be learned (by anyone or anything). To answer these questions, formal models of statistical learning theory are used, requiring math, probability, statistics and optimization.

Course Description Presents the underlying theory behind machine learning in proofs-based format. Answers fundamental questions about what learning means and what can be learned via formal models of statistical learning theory. Analyzes some important classes of machine learning methods. Specific topics may include the PAC framework, VC-dimension and Rademacher complexity. Mostly focuses on binary classification, and almost exclusively focuses on supervised learning and the PAC model.

Principal Topics PAC framework, VC-dimension, Rademacher complexity, bias-variance tradeoffs, no-free-lunch theorems, model selection, boosting, expressive power

Learning Goals/Outcomes The aim of the course is to show students a few possible frameworks in which they can formally analyze machine learning algorithms. Since machine learning is hyped by the media, it is important for students to understand fundamental limits as to what is possible. Furthermore, students will gain an understanding of the assumptions and limitations of the models, and when the theoretical bounds are not useful.

Learning Objectives After taking the course, students will be able to understand the standard frameworks, and analyze standard algorithms within this framework. Students will understand key terms as used in the field, like bias, variance, expressiveness, learnability, training, and generalization. Students will be be able to identify when analysis is possible or impossible. Students will understand that alternative learning models and tasks exist, and know when each model is most appropriate, and in particular be able to evaluate the assumptions and shortcomings of a given model.

Details

We mostly follow the Shalev-Shwartz and Ben-David book, skipping a few chapters and adding a few chapters from Mohri et al.

Classical Statistical Learning Theory We mainly focus on supervised statistical batch learning with a passive learner.

1. Ch 1: Intro to class: what is it about?
2. Ch 2: Formal models (statistical learning), Empirical Risk Minimization (ERM), finite hypothesis class
3. Ch 3: Formal Learning model: Probably-Almost-Correct (PAC)
4. Ch 4: Learning via Uniform Convergence (and concentration inequalities, cf Appendix B and Vershynin)
5. Ch 5: Bias-Complexity Tradeoff[1], no-free-lunch theorems
6. Ch 6: VC-Dimension
7. Ch 26: Rademacher Complexity (and ch 3.1 in Mohri)
8. Ch 27: Covering Numbers

Analysis of Algorithms As time permits, we will analyze standard algorithms.

1. Ch 9: Linear predictors
2. Ch 10: Boosting, AdaBoost

[1] including modern takes, like "Reconciling modern machine-learning practice and the classical bias-variance trade-off," Belkin et al., *PNAS* **116**(32) 2019.

3. Ch 11: Model selection and validation
4. Ch 12: Convex learning problems (generalization bounds)
5. Ch 13: Regularization and Stability
6. Ch 15: Support Vector Machines (SVM)
7. Ch 16: Kernel methods
8. Ch 20: Neural Networks, expressive power, and new results about deep networks (2017–now)[2]

Additional Topics We will cover these as we have time

1. Ch 21: Online Learning
2. Reinforcement learning (ch 17 in Mohri)
3. Background on Information Theory (Appendix E in Mohri)
4. Max Entropy (ch 12 in Mohri)
5. Ch 22: Clustering (K-means, spectral clustering, information bottleneck)
6. Ch 7: Nonuniform Learnability
7. Computational Complexity models (Turing Machines; see Aaronson book)
8. Ch 8: Computational Complexity of learning
9. Ch 14: Stochastic Gradient Descent
10. Recent papers from the literature

Topics in the book that we'll most likely skip Unlikely we'll have time to cover these

1. Ch 17: Multiclass, Ranking and Complex Prediction Problems
2. Ch 18: Decision Trees (VC-dimension)
3. Ch 19: Nearest Neighbors (specialized generalization bounds)
4. Ch 23: Dimensionality Reduction (PCA, Johnson-Lindenstrauss, compressed sensing)
5. Ch 24: Generative Models (parametric density estimation: ML, EM, Naive Bayes, LDA)
6. Ch 25: Feature selection (greedy selection, OMP, LASSO, normalization, autoencoders)
7. Ch 28: Proof of (quantitative version of) Fundamental Theorem of Learning Theory (via Rademacher complexity)
8. Ch 29: Multiclass Learnability (Natarajan dimension, generalization of VC-dimension)
9. Ch 30: Compression bounds (alternative to uniform convergence and stability)
10. Ch 31: PAC-Bayes (putting a prior on the hypothesis)

Supplemental texts and references

A lengthier and up-to-date list will be provided on the course website.

- "Quantum Computing since Democritus," Scott Aaronson, Cambridge University Press (2013).
- "High-Dimensional Probability," by Roman Vershynin, Cambridge University Pres (2018) online version.
- "Introduction to Statistical Learning Theory," Bousquet, Boucheron and Lugosi (2003). 39 pages. Available on the internet, including via SpringerLink.
- "The Nature of Statistical Learning Theory," Vapnik, 2nd edition (2000). Available via SpringerLink.

[2]e.g., " The Expressive Power of Neural Networks: A View from the Width," Lu et al., **NIPS** 2017.

Standard CU-Boulder policies

Accommodation for Disabilities If you qualify for accommodations because of a disability, please submit your accommodation letter from Disability Services to your faculty member in a timely manner so that your needs can be addressed. Disability Services determines accommodations based on documented disabilities in the academic environment. Information on requesting accommodations is located on the Disability Services website. Contact Disability Services at 303-492-8671 or *dsinfo@colorado.edu* for further assistance. If you have a temporary medical condition or injury, see Temporary Medical Conditions under the Students tab on the Disability Services website.

Classroom Behavior Students and faculty each have responsibility for maintaining an appropriate learning environment. Those who fail to adhere to such behavioral standards may be subject to discipline. Professional courtesy and sensitivity are especially important with respect to individuals and topics dealing with race, color, national origin, sex, pregnancy, age, disability, creed, religion, sexual orientation, gender identity, gender expression, veteranstatus, political affiliation or political philosophy. Class rosters are provided to the instructor with the student's legal name. I will gladly honor your request to address you by an alternate name or gender pronoun. Please advise me of this preference early in the semester so that I may make appropriate changes to my records. For more information, see the policies at classroom behavior and the Student Code of Conduct.

Honor Code All students enrolled in a University of Colorado Boulder course are responsible for knowing and adhering to the Honor Code. Violations of the policy may include: plagiarism, cheating, fabrication, lying, bribery, threat, unauthorized access to academic materials, clicker fraud, submitting the same or similar work in more than one course without permission from all course instructors involved, and aiding academic dishonesty. All incidents of academic misconduct will be reported to the Honor Code (*honor@colorado.edu*; 303-492-5550). Students who are found responsible for violating the academic integrity policy will be subject to nonacademic sanctions from the Honor Code Council as well as academic sanctions from the faculty member. Additional information regarding the academic integrity policy can be found at the Honor Code Office website. Students are encouraged to work in groups, however all work turned in must be your own, and you are responsible and accountable for all group work associated with your name.

Sexual Misconduct, Discrimination, Harassment and/or Related Retaliation The University of Colorado Boulder (CU Boulder) is committed to maintaining a positive learning, working, and living environment. CU Boulder will not tolerate acts of sexual misconduct misconduct intimate partner abuse (including dating or domestic violence), stalking, protected-class discrimination or harassment by members of our community. Individuals who believe they have been subject to misconduct or retaliatory actions for reporting a concern should contact the Office of Institutional Equity and Compliance (OIEC) at 303-492-2127 or cureport@colorado.edu. Information about the OIEC, university policies, anonymous reporting, and the campus resources can be found on the OIEC website.

Please know that faculty and instructors have a responsibility to inform OIEC when made aware of incidents of sexual misconduct, discrimination, harassment and/or related retaliation, to ensure that individuals impacted receive information about options for reporting and support resources.

Religious Holidays Campus policy regarding religious observances requires that faculty make every effort to deal reasonably and fairly with all students who, because of religious obligations, have conflicts with scheduled exams, assignments or required attendance. In this class, there should be minimal conflict since there is no attendance grade. If you must miss the final exam due to religious reasons, talk to the professor at the beginning of the semester to make special arrangements. If the homework is due on the date of a religious holiday, you are expected to turn the homework in early. If you have a religious holiday that lasts longer than one week, and so therefore you would not receive the homework with enough time to turn it in early, talk to the professor at the beginning of the semester. See the campus policy regarding religious observances for full details.