CS-GY 9223
Statistical and Computational Foundations of Machine Learning
Spring 2022
Sundays 5:00pm–7:30pm in Jacobs Academic Building, Room 473

Dávid Pál

January 28, 2022

1 Contact

- Instructor: Dávid Pál
- Email: david.pal@nyu.edu
- Virtual Zoom Office: https://nyu.zoom.us/my/dapal
- Ed Discussion Forum (for questions): https://edstem.org/us/courses/20105/discussion/
- Office Hours: Thursdays, 9:30am–11:00am
- Course Webspage: https://david.palenica.com/ml2022/

I do not have an office at NYU. I work from home. I come to NYU only to give lectures. If you need to contact me, please email me and/or arrange a Zoom or Google Hangouts call. Office hours are on Zoom.

2 Course Description

In this course you will learn the theoretical foundations of machine learning. We will explain how a machine learning algorithm can make accurate predictions on unseen data. This property is called *generalization*. We develop a statistical theory of generalization for prediction problems. Along the way we analyze some popular machine learning methods (support vector machines, boosting, perceptron, stochastic gradient descent, decision trees, neural networks, k-nearest neighbor). Besides generalization properties of machine learning algorithms, we will look also at their computational complexity and computational hardness of the underlying optimization problems.

The class will be completely theoretical. In lectures, I will give definitions and prove theorems. In homeworks and exams, I will ask you to solve mathematical problems and prove theorems. You will NOT do any coding, you will NOT do any data analysis, and you will NOT build any machine learning models.

3 Course Prerequisites

You must have taken courses in probability theory, linear algebra, and algorithms. You will benefit from prior exposure to machine learning (ML) (e.g. introductory ML course, neural network course, statistical inference course, practical experience). However, knowledge of ML is not a prerequisite. I will give a brief introduction to ML in the first lecture.

1. **Probability theory.** Random variables, discrete and continuous probability distributions, expectation, variance, correlation, covariance, independence, Bernoulli, binomial and Gaussian distributions, law of large numbers.

2. **Linear algebra.** Vector spaces, subspaces, inner products, Cauchy-Schwarz inequality, norms, triangle inequality, linear functions, matrices, vector-matrix multiplication, matrix multiplication, orthogonal matrices, eigen values, eigen vectors.

3. **Algorithms.** Arrays, linked lists, dictionaries, sorting, searching, time and memory complexity, basics of NP-completeness.

I expect you to know one-dimensional calculus (limits, infimum, supremum, continuity, derivatives and integrals), mathematical notation (set notation, functions, Cartesian product, logical formulas), and elementary combinatorics (combinations, permutations, binomial coeficients). You must be comfortable doing mathematical proofs.

4 Course structure and grading

Homeworks (40% of grade) There will be 4 homework sets. You should spend significant amount of time working on the homeworks. You can work in small groups and/or consult other materials or people. However, in your solution you have to list members of your group, all the materials used and all people you consulted. Type your solutions in LaTeX, print them out, and hand in them in class.

Midterm exam (30% of grade) Midterm exam will take place in place one of the lectures, roughly in the middle of the semester. The questions will have the same format as the the homework problems. You can bring and use any amount of printed materials (notes, books, research papers). However, you cannot use any device connected to the internet (phones, tablets, computers).

Final exam (40% of grade) Final exam will take place during the final period, at the same time as the class. The format of the final exam will be exactly the same as the midterm exam.

Note that the total percentage adds up to 110%. That means you can make up lost points. If you want to do well in the course, make sure to submit solutions to all homework problems.

5 Lectures

There is one 2.5 hour class meeting per week, divided by a 15 minute break. In accordance with NYU guidelines, you must wear a mask covering your face and nose whenever you are attending lecture.

6 Tentative course schedule

I am hoping to cover the topics listed below. Additional topics (decision trees, k-nearest neighbor, neural networks) might be added if time permits.

Week 1	January 30	Generalization error, Test error, Bayes classifier, Hoeffding Bound, Empirical Risk Minimization (ERM), Overfitting, PAC model with finite hypothesis class
Week 2	February 6	PAC model, No-free Lunch Theorem, Vapnik-Chervonenkis dimension, Sauer's lemma, ϵ-nets, ϵ-samples
Week 3	February 13	Upper and lower bounds on the sample complexity for PAC model
Week 4	February 20	Agnostic PAC model, Uniform convergence, McDiarmid's inequality, Rademacher complexity
Week 5	February 27	Upper and lower bounds on the sample complexity for Agnostic PAC model
Week 6	March 6	Online learning model, Halving algorithm, Mistake bound, Perceptron, Winnow, Online-to-batch conversion
Week 7	March 13	**Midterm exam**
Week 8	March 20	**Spring break – No lecture**
Week 9	March 27	Hardness of agnostically learning of halfspaces, Logistic regression, Online convex optimization, Online gradient descent, Stochastic connvex optimization, Stochastic gradient descent
Week 10	April 3	Support Vector Machines (SVM), Primal and dual formulation, Kernels, Hilbert spaces
Week 11	April 10	Margin theory for Support Vector Machines
Week 12	April 17	Weak-learning, Strong-learning, Boosting
Week 13	April 24	TBD
Week 14	May 1	TBD
Week 15	May 8	TBD
Week 16	May 15	**Final exam**

7 Course Policies

7.1 Homeworks

Your solution must be turned **both** by email **and** as a printed out hard-copy in class by the due date. Send emails to david.pal@nyu.edu with subject "Homework #?". Use LaTeX and submit your solution as a PDF file. Use the template provided on the course web page.

Collaboration is allowed on homework, but solutions must be written independently. If you use external sources (books, online documents), list them in your solutions.

Discussion of high-level ideas for solving problems sets or labs is allowed, but all solutions and any code must be written up independently. This reflects the reality that algorithms and machine learning research are rarely done alone. However, any researcher or practitioner must be able to communicate and work through the details of a solution individually.

Students must name any collaborators they discussed problems with at the top of each assignment (list at the top of each problem separately.

Do not write solutions or code in parallel with other students. If problem ideas are discussed, solutions should be written or implemented at a later time, individually. I take this policy very seriously. Do not paraphrase, change variables, or in any way copy another students solution.

7.2 Grade changes

Fair grading is important to me. If you feel that you were incorrectly docked points on an assignment or exam, please let me know - it is very possible I or the TAs misunderstood your work. Do not wait until the end of the semester to bring up grading issues! If you notice something off, please ask ASAP.

7.3 Questions about performance

If you are struggling in the class, contact me as soon as possible so that we can discuss what's not working, and how we can work together to ensure you are mastering the material. It is difficult to address questions about performance at the end of the semester, or after final grades are submitted: by Tandon policy no extra-credit or makeup work can be used to improve a students grade once the semester closes. So if you are worried about your grade, seek help from me early.

7.4 Exams

Midterm and Final will be open-book written exams. They will be 2.5 hours long.

You can bring and use any amount of printed or hand-written material you want (hand-written or printed notes, cheat sheets, textbooks, lecture notes, research papers, etc.) All materials you use have to be in **paper** form. Use of electronic devices will **not** be allowed. This includes computers, laptops, tablets, e-readers, smartphones, smart-watches, headphones, calculators, etc.

Any communication with anybody inside or outside the classroom during the exam will be stricly prohibited. Passing of any materials or objects from one student to another will not be allowed. If you would like to have a cheat sheet or notes from your classmate, make sure to get a copy **before** the exam.

Bring pens, pencils, and extra paper for writing. You can bring ruler and compass if you want, but they will not be necessary or useful. I will bring extra writing paper as well. Calculators of any kind will **not** be allowed. You will have to do all your calculations on paper.

You can bring drinks and food to the exam.

8 Reading

The official textbook and other useful books are listed below. I will post links to research papers, survey articles, lecture notes and other materials on the course web page.

8.1 Books

- **Understading Machine Learning: From Theory to Algorithms** by Shai Shalev-Schwartz and Shai Ben-David [5]. This will be our primary textbook. PDF of the textbook is freely available online at `http://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning`.
- **Foundations of Machine Learning** by Mehryar Mohri, Afshin Rostamizaden, and Ameet Talwalkar [3]. This is a good alternative text. It covers less topics but slightly more in depth.
- **Neural Network Learning: Theoretical Foundations** by Martin Anthony and Peter L. Bartlett [1]. This is a comprehensive book on sample complexity bounds for PAC model, Agnostic PAC model, learning with real-valued losses.
- **An Introduction to Computational Learning Theory** by Michael J. Kearns and Umesh V. Vazirani [2]. This book is a short and easy to read introduction to PAC model.
- **Kernel Methods for Pattern Analysis** by John Shawe-Taylor and Nello Cristianini [6]. This book is a comprehensive introduction to Support Vector Machines (SVM) and other kernel methods.
- **Boosting: Foundations and Algorithms** by Robert E. Schapire and Yoav Freund [4]. This book is about boosting and the theory behind it. The authors are the original inventors of boosting.

9 Additional information

9.1 Moses Center Statement of Disability

If you are student with a disability who is requesting accommodations, please contact New York University's Moses Center for Students with Disabilities (CSD) at 212-998-4980 or mosescsd@nyu.edu. You must be registered with CSD to receive accommodations. Information about the Moses Center can be found at http://www.nyu.edu/csd. The Moses Center is located at 726 Broadway on the 3rd floor.

9.2 NYU School of Engineering Policies and Procedures on Academic Misconduct

Complete Student Code of Conduct here.

A. **Introduction:** The School of Engineering encourages academic excellence in an environment that promotes honesty, integrity, and fairness, and students at the School of Engineering are expected to exhibit those qualities in their academic work. It is through the process of submitting their own work and receiving honest feedback on that work that students may progress academically. Any act of academic dishonesty is seen as an attack upon the School and will not be tolerated. Furthermore, those who breach the School's rules on academic integrity will be sanctioned under this Policy. Students are responsible for familiarizing themselves with the School's Policy on Academic Misconduct.

B. **Definition:** Academic dishonesty may include misrepresentation, deception, dishonesty, or any act of falsification committed by a student to influence a grade or other academic evaluation. Academic dishonesty also includes intentionally damaging the academic work of others or assisting other students in acts of dishonesty. Common examples of academically dishonest behavior include, but are not limited to, the following:

   1. **Cheating:** intentionally using or attempting to use unauthorized notes, books, electronic media, or electronic communications in an exam; talking with fellow students or looking at another person's work during an exam; submitting work prepared in advance for an in-class examination; having someone take an exam for you or taking an exam for someone else; violating other rules governing the administration of examinations.
   2. **Fabrication:** including but not limited to, falsifying experimental data and/or citations.
   3. **Plagiarism:** intentionally or knowingly representing the words or ideas of another as one's own in any academic exercise; failure to attribute direct quotations, paraphrases, or borrowed facts or information.
   4. **Unauthorized collaboration:** working together on work meant to be done individually.
   5. **Duplicating work:** presenting for grading the same work for more than one project or in more than one class, unless express and prior permission has been received from the course instructor(s) or research adviser involved.
   6. **Forgery:** altering any academic document, including, but not limited to, academic records, admissions materials, or medical excuses.

9.3 NYU School of Engineering Policies and Procedures on Excused Absences

Complete policy here.

A. Introduction: An absence can be excused if you have missed no more than 10 days of school. If an illness or special circumstance has caused you to miss more than two weeks of school, please refer to the section labeled Medical Leave of Absence.

B. Students may request special accommodations for an absence to be excused in the following cases:

1. Medical reasons
2. Death in immediate family
3. Personal qualified emergencies (documentation must be provided)
4. Religious Expression or Practice

Deanna Rayment, deanna.rayment@nyu.edu, is the Coordinator of Student Advocacy, Compliance and Student Affairs and handles excused absences. She is located in 5 MTC, LC240C and can assist you should it become necessary.

9.4 NYU School of Engineering Academic Calendar

Complete list here.

The last day of the final exam period is May 17, 2022. Final exam dates for undergraduate courses will not be determined until later in the semester. Final exams for graduate courses will be held on the last day of class during the week of May 11–17, 2022. If you have two final exams at the same time, report the conflict to your professors as soon as possible. Do not make any travel plans until the exam schedule is finalized.

Also, please pay attention to notable dates such as Add/Drop, Withdrawal, etc. For confirmation of dates or further information, please contact Susana: sgarcia@nyu.edu.

References

[1] Martin Anthony and Peter L. Bartlett. *Neural network learning: Theoretical foundations*. Cambridge University Press, 1999.

[2] Michael J. Kearns and Umesh V. Vazirani. *An introduction to computational learning theory*. MIT Press, 1994.

[3] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of machine learning*. MIT Press, 2012.

[4] Robert E. Schapire and Yoav Freund. *Boosting: Foundations and algorithms*. MIT Press, 2012.

[5] Shai Shalev-Schwartz and Shai Ben-David. *Understading Machine Learning: From Theory to Algorithms*. Cambridge University Press, 2014. Available at `http://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning`.

[6] John Shawe-Taylor and Nello Cristianini. *Kernel methods for pattern analysis*. Cambridge University Press, 2004.