DS-GA-1003 Machine Learning (Spring 2022)

Instructor Information

- He He
- Tal Linzen

Course Information

- **Course Description**: The course covers a wide variety of topics in machine learning and statistical modeling. While mathematical methods and theoretical aspects will be covered, the primary goal is to provide students with the tools and principles needed to solve the data science problems found in practice. This course will serve as a foundation of knowledge on which more advanced courses and further independent study can build.
- **Prerequisites**:
    - Introduction to Data Science (DS-GA 1001), or equivalent
    - Statistical and Mathematical Methods (DS-GA 1002), or equivalent
    - **Solid mathematical background**, equivalent to a 1-semester undergraduate course in each of the following: linear algebra, multivariate calculus, probability theory, and statistics (DS-GA 1002 covers the necessary material)
    - **Python programming required** for most homework assignments
    - Recommended: Computer science background up to a course in data structures and algorithms
    - Recommended: At least one advanced, proof-based mathematics course
    - Some prerequisites may be waived with permission of the instructor

Course Goals

Upon Completion of this Course, students will be able to:

- Conduct research in machine learning and its applications
- Pursue advanced studies in machine learning and statistical modeling

Grading

Assignments [40%]

There will be roughly 7-8 homework assignments with both written and programming components. Some homework problems are designated “optional”. These problems will be graded, but will have no effect on the overall homework score.

Exams [60%]

There will be a midterm (30%) in week 7 and a final exam (30%).

Participation

You will be given up to 2% extra credit if you answer other students' questions in a substantial and helpful way. The extra credit may be able to bump up your grades for half a grade (e.g. B -> B+).

Course Schedule

- Week 1 [1/25; Tal]: Introduction, statistical learning theory overview
- Week 2 [2/01; Tal]: loss functions, stochastic gradient descent, subgradient methods
- Week 3 [2/08; Tal]: L1/L2 regularization, Lasso, and Elastic Net
- Week 4 [2/15; He]: SVM
- Week 5 [2/22; He]: Kernel methods
- Week 6 [3/01; He]: Maximum Likelihood, Probabilistic Models
- Week 7 [3/08]: Midterm
    - 3/15: Spring break
- Week 8 [3/22; Tal]: Bayesian methods
- Week 9 [3/29; He]: Multiclass classification, structured prediction
- Week 10 [4/05; Tal]: Trees, Bagging, Boosting
- Week 11 [4/12; He]: Gradient Boosting
- Week 12 [4/19; Tal]: Neural Networks, Backpropagation
- Week 13 [4/26; He]: k-Means, Gaussian Mixture Models, EM
- Week 14 [5/03]: Conclusion

Course Materials

Optional Textbooks & Materials

- Hastie, Tibshirani, Friedman, *Elements of Statistical Learning*, Second Edition, Springer-Verlag, 2009.
- Shalev-Shwartz and Ben-David, *Understanding Machine Learning: From Theory To Algorithms*, 2014.
- David Barber, *Bayesian Reasoning and Machine Learning*, Cambridge University Press, 2012.
- James, Witten, Hastie, Tibshirani, *An Introduction to Statistical Learning*, Springer, 2013.

- Christopher Bishop, *Pattern Recognition and Machine Learning,* Springer, 2007.

Course Policies

Late Assignments

Homework will be accepted for 48 hours after the time it is due, but with a 20% penalty.

Academic Honesty/Plagiarism

The course follows NYU's policy on academic integrity for students.

Disability Disclosure Statement

Students requesting academic accommodations are advised to reach out to the Moses Center for Student Accessibility (CSA) as early as possible in the semester for assistance.

Henry and Lucy Moses Center for Student Accessibility
Telephone: 212-998-4980
Website: https://www.nyu.edu/csa
Email: mosescsd@nyu.edu