GEORGETOWN UNIVERSITY

# Machine Learning: Theory and Application

Fall 2024

https://mathstat.georgetown.edu

| | | | | |
|---|---|---|---|---|
| **Instructor:** | James Schmidt | **Time:** | 1700 – 1815 Tu/Th | **Office Hours:** |
| **Email:** | as5549@georgetown.edu | **Place:** | ICC 107 | By Appointment |

**Description:**

Machine Learning describes a mishmash of computational techniques for "finding patterns in data." The scope of use, analytic tools, algorithms, and results are almost too numerous to meaningfully batch all such applications under a common appellation. Still, we try. This course focuses on *supervised* machine learning, which roughly deals with using historical labeled data to construct a predictor to accurately label future data.

Formally, we work in joint probability space $\mathcal{X} \times \mathcal{Y}$, where $\mathcal{X}$ denotes "input" and $\mathcal{Y}$ denotes "output." While these notions are heuristic, they conveniently describe the scenario wherein data from $\mathcal{X}$ may be readily sampled, whereas data from $\mathcal{Y}$ is difficult or expensive to acquire, and we would like to decision on future use according to how input data $x$ may be associated with label $y$.

In this course, you will learn how to formulate the supervised learning problem in mathematical terms, describe a measure of performance, restrict a search space for constructing models for prediction, optimize performance over the search space, and check for generalization. You will also learn about learning guarantees and of the limits of such results. While grounded in theory, you will gain substantive exposure to implementation through worksheets and programming exercises. Because neural networks are ubiquitous in modern machine learning practice, we will establish foundations for constructing their primary ingredients from which you may confidently navigate use of popular libraries such as PyTorch and TensorFlow.

**Prerequisites:** The following courses or equivalents are strongly recommended. If you would like to take the course but are missing key background, please speak with the instructor.

- Intro Mathematical Statistics (Math 2140)
- Linear Algebra (Math 2250)
- Multivariable Calculus (Math 2370)
- Probability Theory (Math 5051)

**TA:** TBD

**Canvas:** Assignments, data, and starter code will be posted in canvas.

**Useful References:** There is no required text; course notes will be maintained on a git repository and you may find references below useful for supplementary guidance.

**Reference 1.** *https://github.com/schmidttgenstein/math4460.git, course notes.*

**Reference 2.** *Patterns, Predictions, and Actions, M. Hardt and B. Recht*

**Reference 3.** *Neural Networks and Deep Learning, Michael Nielsen*

**Reference 4.** *Real Python (`https://realpython.com/`)*

**Reference 5.** *3Blue1Brown (for probability and neural networks)*

**Reference 6.** *All of Statistics, L. Wasserman*

**Reference 7.** *Foundations of Machine Learning, M. Mohri, A. Rostamizadeh, and A. Talwalkar*

**Reference 8.** *Understanding Machine Learning, S. Shalev-Shwartz and S. Ben-David*

**Reference 9.** *High Dimensional Probability R. Vershynin*

**Workload and Grading:**
The weights, while at the moment nominal, will comprise a genuine probability measure.

1. **Homework 65%**

   There will be roughly ten assignments. They will include problem sets to solidify theoretical concepts and programming assignments to implement learning algorithms.

2. **Exams, 35**
   There will be two midterms and a final exam.

   **Expectations:**

   - Respect: you may and should work together and collaboratively. Ask incisive questions with intellectual curiosity and enthusiasm, challenge assumptions and methods, but do not put down anyone ever.
   - Integrity: personal integrity is one of the most important things you can strive to build up, buttress, and maintain. Take pride in that the honesty of your work reflects who you are.
   - Know University policies

   **My Philosophy:**

   - Your educational experience should be rich, joyful, fun, educational, and challenging. I encourage everyone to have fun learning, while maintaining a rigorous sense of responsibility for actively living up to e.g. the above expectations.
   - I care dearly about both your academic and human success. Please feel welcome to reach out to me with any questions or concerns.

**Tentative Schedule of Topics:**

- Week 1: Introduction
  - Course overview, outline of machine learning, and probability review (ch 1 reference 7, ch 1 reference 8)
- Week 2: Concentration Bounds
  - Markov, Chebyshev, Law of Large Numbers (ch 1 and 2 reference 9)
- Week 3: Detection and Binary Classification
  - Metrics: true/false positive rate, AUC-ROC curves, precision, calibration, equal error rate (ch 2 reference 2)
  - Case study: imbalanced data
- Week 4: Bias / fairness / topical issues in AI/ML + first midterm
- Week 5: Linear Regression
  - Inner product spaces, Hilbert Projection, and Orthogonality Principle (OP) (notes)
  - Regression $\mathbb{E}(y|x)$ as instance of OP and the "geometry of learning"
- Week 6: Beyond Orthogonality Principle
  - Formalizing generic supervised learning problem
  - Restricting the search space: the hypothesis class
  - Empirical risk minimization
- Week 7: Expressivity, Overfitting, and Generalization
  - Fitting to data and its generating source
  - Bias-variance tradeoff
- Week 8: Optimization
  - ML Pipeline in code
  - Gradient descent
- Week 9: Introduction to Neural Networks
  - Universal Approximation Theorem illustrating expressivity
  - Backpropagation instancing gradient descent
- Week 10: Overview of important architectural components
  - Convolutional networks
  - Attention mechanism and transformers
- Week 11: More Optimization
  - Convergence guarantees
  - Stochastic gradient descent
- Week 12-13: Introduction to PAC Learning
  - Concentration bounds revisited: Hoeffding and Glivenko-Cantelli
  - Probably Approximately Correct Learnability
  - VC-dimension and conditions for PAC learnability