Basic Information

Course: EE660: Mathematical Foundations and Methods of Machine Learning
Term: Fall 2024
Instructor: Stephen Tu (stephen.tu@usc.edu)
TAs:
Mohammad Tinati (tinati@usc.edu)
Renyan Sun (renyansu@usc.edu)

Please include [EE660] in the subject of any email related to the course.

Time and Location:

Lecture
Tuesday and Thursday, 4:00-5:50pm, OHE 100D.

Discussion
Friday, 2:00-2:50pm, OHE 100D.

Midterm:
10/8 during class, Location TBD.

Final:
12/12, 4:30pm-6:30pm, Location TBD.

Office Hours:
Monday, 3:00-4:00pm (Stephen), Zoom
Tuesday, 3:00-4:00pm (Stephen), EEB 326.
Wednesday, 5:00pm (Renyan), PHE 320.
Friday, 5:00pm (Mohammad), PHE 320.

Course Prereqs: EE 503, EE 510, and EE 559

Course Description

This course will cover the mathematical foundations of machine learning, including both supervised/unsupervised learning. The focus will be on the mathematical principles underlying modern machine learning algorithms. Topics covered include (but are not limited to) Perceptron, support vector machines, kernel methods, concentration inequalities, empirical process theory, margin theory, VC-dimension, stochastic gradient methods, algorithmic stability, energy-based models, latent variable models, variational autoencoders, diffusion models, model interpretability, model calibration, domain adaptation, and multi-task learning. Students will leave this course with a rigorous understanding of modern machine learning methods.

Textbook

There is no textbook for the course. Instead the reading will be from the course lecture notes.

Note: I will frequently update the lecture notes as the course progresses (filling in new content as we move forward in addition to correcting typos). While I will update the DEN with newer versions of the notes, I recommend instead using the link (https://stephentu.github.io/pdfs/EE660_Lecture_Notes.pdf) and refreshing each time you go to read.

Furthermore, as these notes are a work in progress, please feel free to send me an email if you suspect any typos or errors.

Homeworks

There will be 4 homeworks assignments throughout the semester. These homeworks will mostly be pen and paper exercises, with an occasional coding exercise. For any coding exercises, you are free to use whatever language you choose (although we highly encourage the use of Python and its comprehensive scientific computing ecosystem).

The main deliverable will be a PDF file containing your solutions. **It is heavily encouraged that you use LaTeX to typeset your homework (a sample LaTeX template will be provided before the first homework is assigned)!**

Collaboration Policy: You are free to collaborate with other students on the homework assignments. **However, all solutions must be written by yourself.**

Late Policy: Late submissions will not be accepted, no exceptions! However, we will **automatically drop your lowest score.**

Mini-project

There will also be a mini-project to be completed over the last month of the course. The mini-project will be open-ended, and allow students to apply the material from the course to their research or other interests. The course staff will also provide a few sample project ideas. The project may be more theoretical in nature, e.g., extending a derivation from the course, or applied, e.g., implementing a generative modeling algorithm for a specific problem instance. The main deliverable for the project will be a project report. More details will be provided as the project is assigned.

Grading breakdown

The following table describes how your final grade will be calculated.

Assignment	Weight
Homeworks (4 homeworks in total, we will automatically drop the lowest score).	25%
Mini-project	20%
Midterm	25%
Final	30%

Topics

Supervised Learning

- Perceptron
  - Mistake bound.
  - Generalization bound.
- Support Vector Machines / Kernel Machines
  - SVM primal/dual formulation.
  - Positive definite kernels.
  - Random features.
- Empirical Risk Minimization (short introduction to empirical process theory)
  - Basic concentration inequalities.
  - Loss functions and margin theory.
  - Uniform convergence and Rademacher complexity.
  - VC-dimension and the no free lunch theorem.
- Stochastic gradient methods
  - Basic convergence guarantees.
- Alternative Perspectives on Empirical Risk Minimization
  - Algorithmic stability.
  - Non-uniform priors and compression-based bounds.
- The Curiosities of Overparameterized Models
  - Bias/variance trade-off. Double descent.
  - Optimization is made easier.
  - Which minima generalize?

Probabilistic Modeling

- Energy-based Models
    - Max likelihood training.
    - Langevin dynamics.
    - Score matching.
    - Denoising score matching.
- Normalizing Flows
    - Finite compositions.
    - The log probability ODE.
- Latent Variable Models.
    - Expectation-maximization.
- Variational Autoencoders and GANs
    - Evidence lower bound (ELBO).
    - The GAN min/max loss and its behavior.
- Diffusion Models
    - The DDPM variational inference approach.

Miscellaneous topics

- Uncertainty quantification with conformal prediction.
- Influence functions.
- Domain adaptation.
    - Generalization bounds.
    - Likelihood ratio reweighting.
    - Subspace alignment.
- Multi-task learning.

Schedule

Note: The syllabus will be updated every week with assigned readings corresponding to each lecture. The section(s) in the **Reading and HW** column refer to the lecture notes.

Date	**Topics**	**Reading and HW**
8/27	Intro, binary classification, Perceptron mistake bound.	Sections 1.1, 1.2.1.
8/29	Perceptron risk bound, support vector machines.	Sections 1.2.2, 1.3.1.
9/3	Kernel methods and random features.	**HW1 assigned.** Sections 1.3.2, 1.3.3.
9/5	Empirical risk minimization (ERM) framework, basic concentration inequalities.	Sections 1.4.1, Appendix B.

9/10	Bounds for finite hypothesis classes, hinge and logistic losses, margin theory.	Sections 1.4.3 to 1.4.6.
9/12	Rademacher symmetrization and contraction, generalization bounds.	Sections 1.4.7, 1.4.8.
9/17	Explicit computations of Rademacher complexity.	**HW1 due.** Section 1.4.9.
9/19	VC dimension and examples.	**HW2 assigned.** Sections 1.4.10, 1.4.11.
9/24	PAC learnability and the no free lunch theorem.	Section 1.4.12.
9/26	Stochastic gradient methods.	Section 1.5.
10/1	Algorithmic stability.	Section 1.6.
10/3	Non-uniform priors, compression-based bounds, PAC-Bayes inequalities.	**HW2 due.** Section 1.7.
10/8	**Midterm**	
10/10	**Holiday: Fall Recess**	
10/15	Energy based models, max-likelihood estimation, Langevin dynamics.	**HW3 assigned.** Section 2.1.
10/17	Score matching.	Section 2.2.
10/22	Denoising score matching.	Section 2.3.
10/24	Latent variable models, expectation maximization.	Section 2.4.
10/29	Finish EM, variational autoencoders.	Section 2.5.
10/31	Denoising diffusion probabilistic models.	**HW3 due.** Section 2.6.
11/5	Finish DDPM.	**Mini-project assigned.**
11/7	Normalizing flows, neural ODEs.	Section 2.7.
11/12	Generative adversarial networks.	Section 2.8.

11/14	Principal components analysis.	Section 2.1.
11/19	Uncertainty quantification in ML models via conformal prediction.	Section 3.1.
11/21	Influence functions.	Section 3.2.
11/26	Domain adaptation fundamentals.	Section 3.3.
11/28	**Holiday: Thanksgiving Break**	
12/3	Algorithms for domain adaptation.	-
12/5	Multi-task learning, meta learning.	**Mini-project due.** Sections 3.4, 3.5.