Basic Information

Course: EE660: Mathematical Foundations and Methods of Machine Learning
Term: Spring 2024
Instructor: Stephen Tu (stephen.tu@usc.edu)
TAs:
Fernando Valladares Monteiro (fvallada@usc.edu)
Mutian Zhu (mutianzh@usc.edu)

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

Time and Location:

Lecture
MW 4-5:50pm, OHE 136.

Discussion
F 11-11:50am, OHE 136.

Final:
May 1, 4:30-6:30pm, SGM 123.

Office Hours:
Tu 9-10am (Mutian), PHE 320.
W 3-4pm (Stephen), EEB 326.
F 9:30-10:30am (Fernando), EEB 322.

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) concentration inequalities, empirical process theory, margin theory, VC-dimension, stochastic gradient methods, algorithmic stability, PAC-Bayes inequalities, 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 5 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 will 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.

Grading breakdown

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

Assignment	**Weight**
Homeworks (5 homeworks in total, we will automatically drop the lowest score).	30%
Midterm	30%
Final	40%

Topics

Supervised Learning

- Perceptron
  - Mistake bound.
  - Generalization bound.
- 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.
- Alternative Perspectives on Empirical Risk Minimization
  - Algorithmic Stability.
  - PAC-Bayes Inequalities.
- 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**
1/8	Intro, binary classification, Perceptron mistake bound.	Sections 1.1 and 1.2.1.
1/10	Perceptron risk bound, empirical risk minimization framework.	Sections 1.2.2, 1.3.1, and 1.3.2.
1/15	**Holiday: MLK's Birthday**	
1/17	Sub-Gaussian random variables, Laplace transform method, basic concentration inequalities.	**HW1 assigned.** Section B.
1/22	Bounds for finite hypothesis classes, hinge and logistic losses, margin theory.	Sections 1.3.3 to 1.3.6.
1/24	Rademacher symmetrization and contraction, generalization bounds.	Sections 1.3.7 and 1.3.8.
1/29	Explicit computations of Rademacher complexity.	Section 1.3.9.
1/31	VC dimension and examples.	**HW1 due.** Sections 1.3.10 and 1.3.11.
2/5	PAC learnability and the no free lunch theorem.	**HW2 assigned.** Section 1.3.12.
2/7	Algorithmic stability framework, stability of SGD.	Section 1.4.
2/12	Finish stability analysis of SGD, stability of Gibbs-ERM.	
2/14	PAC-Bayes inequalities and various examples.	Section 1.5.
2/19	**Holiday: President's Day**	
2/21	Curiosities of modern deep networks.	**HW2 due, HW3 assigned.**

2/26	Energy based models, max-likelihood estimation, Langevin dynamics.	Section 2.1.
2/28	Score matching.	Section 2.2.
3/4	Denoising score matching, midterm review.	**HW3 due.** Section 2.3.
3/6	**Midterm (GFS 116)**	
3/11	**Holiday: Spring Recess**	
3/13	**Holiday: Spring Recess**	
3/18	Latent variable models, expectation maximization.	**HW4 assigned.** Section 2.4.
3/20	Finish EM, variational autoencoders.	Section 2.5.
3/25	Denoising diffusion probabilistic models.	Section 2.6.
3/27	Finish DDPM.	
4/1	Normalizing flows, neural ODEs.	**HW4 due.** Section 2.7.
4/3	Generative adversarial networks.	**HW5 assigned.** Section 2.8.
4/8	Uncertainty quantification in ML models via conformal prediction.	Section 3.1.
4/10	Influence functions.	Section 3.2.
4/15	Domain adaptation fundamentals.	Section 3.3.
4/17	Algorithms for domain adaptation.	
4/22	Multi-task learning, meta learning.	Sections 3.4 and 3.5.
4/24	In class office hours.	**HW5 due on 4/28.**