ESE 546: Principles of Deep Learning (Fall 2020)

Instructor
Pratik Chaudhari (pratikac@seas.upenn.edu)
Website: https://pratikac.github.io

Teaching Assistants
Evangelos Chatzipantazis (vaghat@seas.upenn.edu), Rahul Ramesh (rahulram@seas.upenn.edu), Dushyant Sahoo (sadu@seas.upenn.edu), Ronguang Wang (rgw@seas.upenn.edu), Shiyun Xu (shiyunxu@seas.upenn.edu)

Canvas
https://canvas.upenn.edu/courses/1528989

Piazza
https://piazza.com/upenn/fall2020/ese546

Course Description
Deep networks are at the heart of modern approaches in computer vision, natural language processing and robotics. Design of these networks requires a combination of intuition, theoretical foundation and empirical experience; this course discusses general principles of deep learning that cut across these three. It develops insight into popular empirical practices with a focus on the training of deep networks, builds the theoretical skills to develop new ideas in deep learning and to deploy deep networks in real world applications. A fair degree of mathematical and programming proficiency is necessary to complete the coursework.

You can go through the notes for the Fall 2019 offering at https://pratikac.github.io/pub/19_ese546.pdf. The content will undergo some changes in Fall 2020 (see the syllabus below) and the course will be a bit less theoretical.

Prerequisites
Required

1. Proficiency in programming (ENGR 105, CIS 110, CIS 120 or equivalent). All assignments will be based on Python but if you have used a similar language like MATLAB before, you should be able to pick up Python easily. Recitation sessions will provide preparatory material.
2. Probability (ESE 301, STAT 430, CIS 261, ENM 503, ESE 530 or equivalent)
3. Linear Algebra (Math 312, EAS 205 or equivalent)

Recommended

1. Machine Learning or Data Analysis (ESE 305, ESE 402, ESE 545, CIS 519/520, or a first machine learning course)
2. Optimization (ESE 304, ESE 504, ESE 605)

Evaluation

- 4 bi-weekly homeworks (10% each)
- Mid-term exam (20%)
- Final exam (20%)
- Course project (15%)
- You will write a summary (it can be as elaborate as you like but at least 2 pages) that demonstrates your understanding of the material *in your own words* for each of the 4 modules in the course. These summaries will together make up for 5% of your final grade. There is no partial credit here, depending on the quality of your summary, you either get all the 5% or none.

Each student will have 10 “late days” to use during the semester which can be used to submit the deliverables later than they are due. Deliverables that are submitted late after exhausting this quota will result in 50% credit deduction per day.

Textbook

Detailed instructor notes will be provided for the course and they will form the primary reference/textbook. Reading material for every lecture will consist of recently published research papers and excerpts from the following books.

1. Pattern Recognition and Machine Learning by Christopher Bishop. You can find a PDF here.
2. Deep Learning by Ian Goodfellow, Yoshua Bengio and Aaron Courville. https://www.deeplearningbook.org
3. Dive into Deep Learning by Aston Zhang, Zack Lipton, Mu Li and Alex Smola available at https://d2l.ai is a good reference for more applied parts of the course.

There are a number of other textbooks and reading materials that will be suggested in addition to these in the lectures. In particular, the following books are intended as advanced reference material. We will not use them directly in the course but you are encouraged to go through them (in this order) to improve your understanding.

1. Machine Learning: A Probabilistic Perspective, by Kevin P. Murphy. You can find a PDF here.
2. The Elements of Statistical Learning by Trevor Hastie, Robert Tibshirani, Jerome Friedman. You can find a PDF here.
3. Information Theory, Inference, and Learning Algorithms by David MacKay. You can find a PDF here.

Computational Resources

We will use PyTorch (https://pytorch.org) for homeworks in the course; you can certainly use other libraries such as TensorFlow (http://tensorflow.org) but support will be limited. Most homework problems will be such that you can solve them on your laptop. For other problems, you can use the following.

1. Free: Google Colab (https://colab.research.google.com) is a very good platform to use for doing your homework. Gradient (https://gradient.paperspace.com) is another free tool with more generous compute resources (6-hour timeouts and persistent sessions). If you haven't used it already Google Cloud Project gives $300 of starter credits (https://cloud.google.com/free); this should be plenty to do the homework for this course.
2. Paid: You can also sign up for Google Colab Pro (https://colab.research.google.com/signup) for a very reasonable $10/month which gives you access to faster GPUs and less restrictive pre-emption of jobs.

Every enrolled student will also get some Amazon AWS cloud credits ($100) for doing the homeworks and the project; recitations will cover tips on how to make the best use of these credits.

Course Syllabus

Module 1: Introduction to deep networks (8 lectures)

Background on linear regression, kernel methods

Backpropagation, neural architectures (convolutions, recurrent networks, attention modules), loss functions, data augmentation, regularization techniques

Module 2: Training deep networks (6 lectures)

Background on convexity, bias-variance trade-off

SGD, accelerated variants of SGD, early-stopping/implicit regularization, uniform convergence-based generalization bounds

Module 3: Energy landscape of neural networks (7 lectures)

Background on principal component analysis

Deep linear networks, digging deeper into two-layer neural networks, gradient flows

Module 4: Variational inference and generative models (7 lectures)

Background on information theory

Auto-encoders, ELBO, Bayesian neural networks, Generative Adversarial Networks (GANs)

Lecture	Date	Day	Topic	Notes
1	9/2	W	Introduction, History of deep learning	HW0 out (not graded)

Rec 1	9/4	F	Python, Numpy, Google Colab; ngork, mounting Google Drive, wand.db	Rahul
	9/7	M	Labor Day	
2	9/9	W	Linear Regression, Perceptron, stochastic gradient descent	HW0 due, HW1 out
Rec 2	9/11	F	PyTorch I: Syntax, basics of autograd, various modules, layout of the library	Ronguang
3	9/14	M	Kernel methods	Course Selection Period ends on 9/15
4	9/16	W	Beginning of neural networks	
Rec 3	9/18	F	PyTorch II	Shiyun
5	9/21	M	Backpropagation	
6	9/23	W	Convolutional architectures	
Rec 4	9/25	F	Using the AWS cloud	Rahul
7	9/28	M	Data Augmentation, Loss functions	HW1 due on 9/29, HW2 out
8	9/30	W	Dropout, Batch-Normalization	
Rec 5	10/2	F	Neural architectures: hall of fame	Dushyant
9	10/5	M	Recurrent, Attention-based architectures	
			Module 2 begins	
10	10/7	W	Background: Optimization (convexity, gradient descent)	Module 1 Summary due
Rec 6	10/9	F	Recap of convexity	Shiyun
11	10/12	M	Background: Generalization (bias-variance tradeoff, double descent)	HW2 due, Drop Period ends on 10/12
12	10/14	W	Stochastic gradient descent (SGD) I	
Rec 7	10/16	F	Tricks of the trade: Training deep networks with SGD	Ronguang
13	10/19	M	SGD II: Momentum (heavy-ball, Nesterov), Adam	
14	10/21	W	Early stopping and implicit regularization	Midterm, Project proposal out
Rec 8	10/23	F	Recap of regularization, generalization bounds	Evangelos, Module 2 Summary due

15	10/26	M	Generalization bound using uniform convergence	
			Module 3 begins	
16	10/28	W	Background: Principal Component Analysis	Project proposal due, HW3 out
Rec 9	10/30	F	Vignette: Object Detection	Ronguang Grade Type Change deadline is 10/30
17	11/2	M	Deep Linear networks	
18	11/4	W	Energy landscape of two-layer networks	
Rec 10	11/6	F	Vignette: NLP Architectures	Dushyant
19	11/9	M	Gradient flows	Last date to withdraw from course is 11/9
20	11/11	W	SGD as a Markov chain	HW3 due, HW4 out
Rec 11	11/13	F	Vignette: Deep reinforcement learning	Rahul
21	11/16	M	Where does SGD converge? Entropy-SGD	
			Module 4 begins	
22	11/18	W	Background: Information theory	Module 3 Summary due
Rec 12	11/20	F	Recap of SGD, Info theory	Evangelos
23	11/23	M	Variational Inference	
Rec 13	11/25	W/F	Vignette: Meta-Learning	Pratik HW4 due, Friday schedule
	11/27	F	Thanksgiving Break	
24	11/30	M	Auto-Encoders, ELBO I	
25	12/2	W	ELBO II	
Rec 14	12/4	F	Hands-on variational auto-encoders	Shiyun
26	12/7	M	Bayesian neural networks	
27	12/9	W	Generative Adversarial Networks	Project due
28	12/10	R/M	Recap of post-midterm topics, project presentations	Monday schedule, Module 4 Summary due
	12/15-12/22		Final Exam	

Academic Integrity

You are encouraged to collaborate with your peers for solving problems in the homework, reading books and curating other instructional materials to improve your understanding of the concepts taught in the class. While doing so, you might generate code/pseudo-code/solutions for the homeworks/project. When you begin to write your submission you should keep aside all these materials (including your friends) and do things from "from scratch". In short, everything you write/code and submit should be your own work done independently.

You should disclose all collaborations in your submission at the top. If you came across some code as a part of your homework/project you must mention it.

Collaboration is different from cheating. The latter will have serious consequences. Cheating is defined as attempting, abetting or using unauthorized assistance (knowledgeable friend who is not taking the class) or material (e.g., online code). Some examples of cheating are: copying someone else's work for homework/exams, handing in someone else's work as your own, handing in stuff from the Internet as your homework. This will not be tolerated. Your score for that particular homework or exam will be zeroed out if found guilty, *and* this incident will be reported to the university.

Inclusive Environment

If you have any special needs that affects your coursework/attendance, please talk to the instructor and I will do my best to accommodate them. We are all adults here and we should together create an open and inclusive academic environment. This includes being respectful while interacting with your peers, being receptive to their ideas/thoughts, and, in general, maintaining a scholarly discourse both inside and outside the classroom. Transgressions will be reported to the university.