CS281: Advanced Machine Learning (Fall 2018)

Instructor:	Prof. Alexander (Sasha) Rush
Teaching Fellows:	Anirudh Suresh, Carl Denton, Yaniv Yacoby, Isaac Lage, Demi Guo, and Luke Melas-Kyriazi
Lectures:	Monday and Wednesday, 1:30-2:45pm
Location:	Science Center D
Section:	TBD
URL:	`http://cs281.fas.harvard.edu/`

This Syllabus

This document reviews the course description, prerequisites, major components of the course, and course policies. All information about the lectures, sections, office hours, assignments, textbook, readings, course grading, and Piazza discussion board may also be found on the course website.

TL;DR

Advanced statistical machine learning and probabilistic data analysis. Topics include: Variational inference, graphical models, deep learning, text modeling, unsupervised learning, dimensionality reduction and visualization.

Course Description

This course is about learning to extract statistical structure from data, for making decisions and predictions. The course will cover many of the most important mathematical and computational tools for probabilistic modeling, as well as examine specific models from the literature and examine how they can be used for particular types of data.

In addition to the mathematical fundamentals. There will be a heavy emphasis on implementation and computational aspects of inference. For each model and inference technique studied there will be accompanying demos and exercises asking students to implement and apply to real data.

A major theme of the class will be inference in the face of rich, highly parameterized models. In recent years many successful machine learning approaches have towards complicated non-linear models with millions-to-billions of parameters. While the first part of the course will focus on the basics of well-understood inference approaches (multivariate normals etc.), the course will quickly move to cover recent research on more difficult settings.

Objectives

This course has the following aims:

- Use the language of probabilistic modeling to describe and represent real-world problems.
- Understand different ML algorithm in terms of the underlying inference challenges.
- Implement a variety of inference techniques in simple and declarative ways.
- Seamlessly combine non-linear methods, e.g. neural networks, into generative models.

Prerequisites

You should feel comfortable with probability: joint densities, conditional distributions, etc. An undergraduate level is fine — there will not be any measure theory and such background is not required. There will be linear algebra, e.g., solving linear systems and thinking about eigenvalues/eigenvectors. We will regularly use multivariable calculus. Additionally, prior exposure to machine learning on the level of CS181 is highly recommended.

This course additionally has a significant coding component. We expect intermediate knowledge of Python and Numpy (or similar). The course will make heavy use of the PyTorch library. All of the homework assignments will require some amount of coding, and some projects will require management of longer-running computational experiments.

Readings and Lecture Preparation

Each week will have some required reading, typically from the Murphy or Bishop text and some occasional papers. There will be a strongly-recommended weekly section led by the teaching staff. The material covered in section will provide examples and applications of the material presented in lecture and will be pertinent to the assignments and to the final projects; it is highly recommended that you attend section.

Assignments

There will be six assignments due over the course of the term. Please see the course website for the homework schedule.

Homeworks will consist of a mix of mathematical and coding problems. To emphasize the continuity between these two parts, this year we will be experimenting with a more fully integrated use of Jupyter notebooks. All homeworks will be distributed as interactive notebooks through Google Colab, a GPU accelerated notebook service. Since Jupyter includes LaTeX support, problems will alternate between the two types of exercises. Homeworks can be completed entirely through this web-service, or completed offline.

The assignments themselves should be turned in as Jupyter or as LaTeX PDF's. All assignments should be submitted electronically by 5:00pm on the due date via Canvas. It is your responsibility to ensure that the PDF files are readable.

Collaboration Policy You are encouraged to discuss high-level solutions and problem-solving approaches with other students in the class, but ultimately you must write your own code and produce your own results. If you have collaborated with other students in the planning and design of solutions to the assignments, provide their names on your writeup in the space provided.

Self-Grading Policy Once the submission deadline has passed, students will receive the answer key for the assignment. Students will then be responsible for using this answer key to submit their own scores; this process is there to ensure that students check their understanding and to provide very rapid feedback. The staff will spot-check submitted scores for accuracy. *It is a honor code violation to look at the answer key if you haven't yet turned in your assignment (e.g. using late days). It is also an honor code violation to look at an answer key from a previous iteration of the class.*

Regrading Policy If you think there is a mistake with the grading of your midterm or one of your assignments, you may submit it to the course staff for a regrade. Note that this regrade is a new draw from the grading distribution and may result in your grade moving up or down. There is a time limit on these requests: you may only ask for a regrade within two weeks of the section at which the assignment/midterm was returned.

Late Policy You have a total of 5 late days that can be used for assignments. Weekends count as late days. Please use your late days carefully. Late days cannot be applied to the final project write-up.

Midterm Exams

There will be two in-class midterm exams, scheduled for October 24, 2018 and December 5, 2018. This first exam will cover material from the beginning of the class, through the October 15th lecture on Variational Inference. The second exam will cover material from undirected graphical models through the end of sampling.

Grading

Please see the Grading section of the course website for more details.

Textbooks and References

The course will mainly follow the comprehensive textbook:

Machine Learning: A Probabilistic Perspective Kevin P. Murphy, MIT Press, 2012.

Some students may already own (Bishop), which often has a more focused, and some might say, clearer presentation of the topics:

Pattern Recognition and Machine Learning Christopher M. Bishop, Springer, 2006.

Either is sufficient for the course and we will provide references to both texts throughout.

These are other (free online!) books on machine learning and related topics that you may find helpful, but that are completely optional.:

Information Theory, Inference, and Learning Algorithms David J.C. MacKay, Cambridge University Press, 2003. Freely available online at `http://www.inference.phy.cam.ac.uk/mackay/itila/`. A very well-written book with excellent explanations of many machine learning topics.

Bayesian Reasoning and Machine Learning David Barber, Cambridge University Press, 2012. Freely available online at `http://web4.cs.ucl.ac.uk/staff/D.Barber/pmwiki/pmwiki.php?n=Brml.Online`.

The Elements of Statistical Learning Trevor Hastie, Robert Tibshirani, and Jerome Friedman, Springer, 2009. Freely available online at `http://www-stat.stanford.edu/~tibs/ElemStatLearn/`.

These are books on some specialized topics that you may find useful:

Gaussian Processes for Machine Learning Carl Edward Rasmussen and Christopher K.I. Williams, MIT Press, 2006. Freely available online at `http://www.gaussianprocess.org/gpml/`.

Non-Uniform Random Variate Generation Luc Devroye, Springer-Verlag, 1986. Freely available online at `http://luc.devroye.org/rnbookindex.html`.

Probabilistic Graphical Models: Principles and Techniques Daphne Koller and Nir Friedman, MIT Press, 2009.

Numerical Optimization Jorge Nocedal and Stephen J. Wright, Springer, 2006.

Bayesian Data Analysis Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. CRC, 2013.

Elements of Information Theory Thomas M. Cover and Joy A. Thomas, Wiley, 1991.

Monte Carlo Statistical Methods Christian P. Robert and George Casella, Springer, 2005.

Piazza

Most questions about the course, lecture or section material, or the assignments should be addressed via

Piazza at `https://piazza.com/harvard/fall2018/cs281`. The course instructors will check this discussion board and make an effort to post responses within a day. Students taking the class are also encouraged to post responses. Code examples can be posted, but don't post anything you wouldn't be expected to share with other students in the class as per the collaboration policy. Long, detailed questions are probably best answered during office hours. Questions that are not appropriate for the discussion board may be sent to instructors privately through Piazza. Use your judgement. **Note that setting a public post as anonymous will hide your name from classmates. Instructors will always see your name on a public, anonymous public, or private post.**