STAT 991: Topics in Deep Learning

Department of Statistics, The Wharton School, University of Pennsylvania, Spring 2019

Time:	Thursdays 12:00pm-1:20pm	**Place:**	Huntsman Hall F38

Objectives: This advanced seminar course will explore several topics in deep learning. We will discuss both theory and applications.

Students will work in groups of 2-3 people and present a topic over several lectures. The presentations will summarize either the basic foundations of the area or the work of several research papers on a topic. They will include necessary background, algorithms, in-class code demonstrations, as well as results and proofs (in case of theory). Finally, the goal of the course will also be to identity new research directions.

I am hoping that this course will provide a venue for discussion for students interesed in deep learning and related areas at Penn.

Prerequisites: You are expected to have some basic familiarity with deep learning. In particular, you should know what the following terms mean: deep net, hidden layer, activation, weight, bias, ReLU, backpropagation, SGD, dropout, batchnorm, CNN, RNN, Keras. In addition, you are expected to have a basic familiarity with statistics and machine learning.

These have been covered in STAT 991 during Fall 2018, and the lecture notes for that course (provided on Canvas) are one way to learn the prerequisites. For the remainder, you are expected to learn it yourself. Good resources include:

1. Konrad Kording, Jeffrey Cheng, Lyle Ungar's class, CIS 700-004 on M/W from 4:30-6pm. Students can apply for a permit here: `https://forms.cis.upenn.edu/waitlist/index.php`
2. Andrew Ng's course, (lecture videos)

   `https://www.coursera.org/learn/neural-networks-deep-learning`.

Topics: A list of possible topics are included below. We can also cover other topics based on student interest.

A few references and papers for each topic have been posted on Canvas, under /Files/Papers. These are a starting point for each area. You may find many more resources on the web (Google search and Scholar).

- **Sequential decision-making: from bandits to deep reinforcement learning**

  Topics: Bandits, contextual bandits, regret bounds, RL, robust control, online learning, Bayesian optimization, applications

  In addition to the materials posted on Canvas, the following are recommended:

  1. Peter Bartlett's course,

     `https://www.stat.berkeley.edu/~bartlett/courses/2014fall-cs294stat260/`.

2. Ambuj Tewari and Susan Murphy's course,
   http://dept.stat.lsa.umich.edu/~tewaria/teaching/STATS710-Fall2016/
3. How is Google using bandits in its analytics engine?
   https://support.google.com/analytics/answer/2844870?hl=en&authuser=0
4. Rob Schapire's talk on contextual bandits,
   https://www.youtube.com/watch?v=N5x48g2sp8M.
5. David Silver's course, (lecture videos and notes)
   http://www0.cs.ucl.ac.uk/staff/d.silver/web/Teaching.html.
6. Readable blog post on RL,
   https://lilianweng.github.io/lil-log/2018/02/19/a-long-peek-into-reinforcement-learning.html
7. Sergey Levine's Deep RL course,
   http://rail.eecs.berkeley.edu/deeprlcourse/
8. Ben van Roy's RL course (more mathematical)
   https://web.stanford.edu/class/msande338/index.html
9. Software:
   Open AI Gym tutorial,
   https://blog.openai.com/spinning-up-in-deep-rl/,
   various bandit libraries,
   https://github.com/bgalbraith/bandits,
   https://github.com/alexrutar/banditvis,
   https://github.com/mpatacchiola/dissecting-reinforcement-learning,
   https://github.com/david-cortes/contextualbandits

- **Scalable deep learning**: how to make deep learning scale up to truly massive datasets, or how to make it be fast enough to be run on edge devices such as cellphones

  Topics: parallel computing/programming (MPI), communication theory, distributed systems, ML and systems, MapReduce/Spark/GraphLab, distributed algorithms, cloud computing

  In addition to the materials posted on Canvas, the following are recommended:

  1. Tutorials on parallel computing and MPI.
     https://computing.llnl.gov/tutorials/parallel_comp/,
     https://computing.llnl.gov/tutorials/mpi/
  2. Reza Zadeh's course Distributed Algorithms and Optimization
     https://stanford.edu/~rezab/dao/
  3. Alex Smola's course on scalable ML,
     http://alex.smola.org/teaching/berkeley2012/index.html.
  4. Readable blog post on distributed DL.
     https://blog.skymind.ai/distributed-deep-learning-part-1-an-introduction-to-distributed-tr

- Uncertainty quantification: How can we evaluate the certainty in our predictions with complex deep networks?

  Topics: Bayesian statistics, variational inference

  In addition to the materials posted on Canvas, the following are recommended:

  1. UAI Workshop: https://sites.google.com/view/udl2018/schedule?authuser=0

- Generative models

  Topics: GANs, non-adversarial models

- Others: Data augmentation, Applications (medical, climate), Automated ML,

Computation: There is no formal computational component of the course. However, it can be extremely helpful to build experience doing deep learning yourself. For this, if you are coming from a statistics background, the easiest route is to use Keras in R. If you have experience using other computational frameworks, you may use those in your presentation.

In the Fall 2018 edition of the class, we have some small deep learning experiments in class. The code for these is provided on Canvas under the folder /Files/Code/ for your conveniance. To set up your environment, follow the following steps:

1. To prepare, you need to install the Python distribution Anaconda on your laptop. For this there are several tutorials online.
2. You also need to install Keras: https://github.com/keras-team/keras. The code to do this is included is included in the R scripts posted on Canvas.
3. We use material from the book Deep Learning in R, by Chollet and Allaire. See https://github.com/jjallaire/deep-learning-with-r-notebooks.

Instructor: Edgar Dobriban, dobriban@wharton.upenn.edu, Office: 465 JMHH. Office Hours: by appointment

Course Page: Canvas, for announcements and materials: https://canvas.upenn.edu/courses/1440266

Feedback: I am interested to hear about your experience and suggestions for the class.

Grading Policy: The course grade will be driven by two factors: presentation (80%), and class participation (20%). The components of each are

- Presentation: Clarity (ability of others to follow). Correctness. Coverage (did you cover the important parts). Insight.

  Think of the presentation as a course project. You will need to prepare the presentation (slides to be presented in class, or cca 6 pages of lecture notes) and provide it to the class 24 hours in advance. These will be posted on Canvas.

  There will be a Google Sheet where you can sign up for specific topics and dates. I can guide with choosing topics, literature search, and structuring the presentation.

- Class participation: Attendance, asking questions.