Advanced Machine Learning (GR5242)
Fall 2018

John P. Cunningham, Peter Orbanz
jpc2181, po2197
Ian Kinsella, Gabriel Loaiza-Ganem, Andrew Davison, JinHyung Lee
iak2119, gl2480, ad3395, jl4303

Course Syllabus

Description

The third course in the Machine Learning sequence, culminating the skills and knowledge from Statistical Computing & Introduction to Data Science (GR5206) and Statistical Machine Learning (GR5241). The course will go into depth on probabilistic models and deep learning, and will include homeworks and a substantial course project in the `python` language.

Administrative

Lectures

- Section 001:
  Monday and Wednesday, 4:10PM–5:25PM
  Location: 501 Northwest Corner Building
- Section 002:
  Tuesday and Thursday, 4:10PM–5:25PM
  Location: 614 Schermerhorn Hall

Instructors

- John Cunningham
  Email: jpc2181@columbia.edu
- Peter Orbanz
  Email: po2197@columbia.edu

Teaching Assistants

- Ian Kinsella, Gabriel Loaiza-Ganem, Andrew Davison, JinHyung Lee
  Office Hours:
  - Mondays 5:30pm-7:00pm, Room 1025, Department of Statistics
  - Tuesdays 5:30pm-7:00pm, Room 1025, Department of Statistics

  Office: Department of Statistics, 10th Floor School of Social Work, 1255 Amsterdam
  Email: {iak2119, gl2480, ad3395, jl4303}@columbia.edu

Email and Piazza Guidelines

- Email Policy:
  All questions pertaining to the course (organization, curriculum, material, etc.) should be asked on Piazza, where other students will be able to benefit from the answers. Questions containing sensitive and/or personal information can be directed to the TA's who will redirect them to the professors as appropriate. Please do not send emails directly to the professors.

- Piazza Policy:
  The TA's will manage a Piazza forum for students to ask, and help answer, questions pertaining to course organization and material. Students are highly encouraged to browse previously answered questions and contribute answers to those still awaiting responses. Questions must be specific, self contained, and answerable with a single response. For example, a question troubleshooting the performance of a model needs to provide details about the model architecture, training procedure, dataset, and performance metrics. If your question does not fit well within these confines, it should be brought to office hours where that type of discussion is more appropriate. Furthermore, the Piazza forums and office hours are specifically intended for discussion about the models and concepts covered in the course, not general Python or Tensorflow programming/debugging questions. Questions pertaining to these topics, for example "Why do I get this error message when I run my code?", should be directed to appropriate online resources such as Stack Overflow.

Course Resources

- Website (primary source for course information): `https://www.adavison.co.uk/teaching/AdvancedML18/`
- Courseworks (for homework submission, etc): `https://courseworks2.columbia.edu/`

Prerequisites

- GR5206 and GR5241

Grading and Academic Integrity

We take the honor code very seriously; students caught cheating or otherwise in violation will face disciplinary action. Please note:

`http://barnard.edu/node/2875`
`https://www.college.columbia.edu/academics/academicintegrity`

Your grade will be determined by three different components:

- **Homework (20%).** Homework will contain both written and `python` data analysis elements. This is due online at `https://courseworks2.columbia.edu` on Wednesdays at 4:10pm. **Your two lowest homework grade will be automatically dropped at the end of the term.**
- **Midterm Exam (40%).** This will be given in class during midterm week. **Midterm dates: October 18 Thursday (Section 002) and October 22 Monday (Section 001).**
- **Final Project (40%).** A substantial course project will be executed and submitted by teams of 1-4 students. **Final Project due date: December 10 Monday (both sections).**

Failure to complete any of these components may result in a D or F.

Late Work and Regrading Policy: No late work or requests for regrades are accepted. To accommodate unexpected circumstances, we have implemented two important features:

- Your lowest two homework grades will be automatically dropped at the end of the term.
- You may submit and resubmit your homework as many times as you like up until the deadline. This means that you should submit any partial solutions as you complete them, to make sure you receive as much credit as possible for the work you have done. After the deadline, the system will not allow you to submit your homework. If you do not submit anything by the deadline, you will get a 0. **There will be no exceptions to this rule. Submit your homework early.**

Reading Material

No explicit readings will be assigned. Rather, students should use the following two books as supporting references.

- Bishop, C. *Pattern Recognition and Machine Learning.* Springer-Verlag, 2006.
- Murphy, K. *Machine Learning: a Probabilistic Perspective.* MIT Press, 2012.

Topics

The course will be taught in two halves, with the following topics:

- Part I: Probabilistic models (Orbanz)
  - Markov and hidden markov models
  - Graphical models
  - Sampling and MCMC algorithms
  - Variational inference
  - Neural network basics
- Part II: Deep learning (Cunningham)
  - NN software
  - Neural networks: convolutional, recurrent, etc.
  - Reinforcement learning
  - Dimension reduction and autoencoders