Gaussian Processes and Kernel Methods (G8325)

Fall 2015
https://courseworks.columbia.edu

John P. Cunningham
jpc2181@columbia.edu

Course Syllabus

Description

Kernels – which implicitly achieve rich feature space representations – are one of the most widely deployed tools in machine learning, spanning much of the field: from supervised to unsupervised learning, classification to regression, frequentist to Bayesian, etc. One particularly fruitful application of this general class of techniques is the use of gaussian processes for nonparametric regression. This course will explore the cutting edge of the modern literature on gaussian processes, covering their basic use, the theory underlying them, kernel choices, speed and scaling issues and techniques, approximate inference, their use for optimization and active learning, and more. We will also spend some weeks covering non-probabilistic kernel techniques, including the basic theories of reproducing kernel hilbert spaces, kernel mean embeddings, kernel two sample tests, and more.

This is a PhD-level course, and its focal point will be a substantial project by the student (or group of two students). Students have substantial freedom in choosing these projects, from creating novel research-grade methods, to contributing to open source machine learning projects, to analyzing data of interest, to exploring a theoretical topic. Real data analysis problems and problems that connect to the student's ongoing research agenda are particularly welcome.

Administrative

Lecture

- Mondays, 10:10AM–12:00PM
  Location: TBD

Instructor

- John Cunningham
  Office: Department of Statistics, Room 1026, 10th Floor School of Social Work, 1255 Amsterdam
  Email: jpc2181@columbia.edu

Office Hours

- Individual meetings regarding course projects will be scheduled by appointment at the appropriate time.

Prerequisites

- Stochastic processes to the level of a basic understanding of gaussian processes
- Statistical machine learning such as the topics covered in W4400, W4240, or similar.
- Probability, statistics, linear algebra, and a modest understanding of convex optimization
- Programming skills

Grading and Academic Integrity

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

> "We... resolve to uphold the honor of the College by refraining from every form of dishonesty in our academic life. We consider it dishonest to ask for, give, or receive help in examinations or quizzes, to use any papers or books not authorized by the instructor in examinations, or to present oral work or written work which is not entirely our own, unless otherwise approved by the instructor.... We pledge to do all that is in our power to create a spirit of honesty and honor for its own sake."
> `http://barnard.edu/node/2875`
> `https://www.college.columbia.edu/academics/academicintegrity`

Your grade will be determined by three different components:

- **Homework (10%).** Two or three homework sets will be given to ensure students are keeping pace. Homework will contain both written and programming/data analysis elements.
- **Attendance and Participation (40%).** This will be given in class during midterm week. You will be permitted use one handwritten page, front and back, of notes.
- **Course Project (50%).** This will be given in class during the finals period. You will be permitted use one handwritten page, front and back, of notes.

Failure to complete any of the first three components may result in a D or F.

Late Work and Regrading Policy: No late work or requests for regrades are accepted. Drafts of homeworks may be submitted to courseworks ahead of time, and you are strongly encouraged to do so.

Course Project: The course projects will be the focus of the latter half of this course. Projects can take a variety of forms, from contributing to open source machine learning projects, to analyzing data of interest, to advancing a theoretical topic. We will spend substantial time developing ideas for projects, tracking and discussing progress, and presenting final work product. Individual projects are ideal, though projects with groups of two may also be appropriate.

Homework: Students are encouraged to work together, but homework write-ups must be done individually and must be entirely the author's own work. Homework is due at the **beginning** of the class for which it is due. **Late homework will not be accepted under any circumstances.** To receive full credit, students must thoroughly explain how they arrived at their solutions and include the following information on their homeworks: name, UNI, homework number (e.g., HW02), and class (STAT G8325). All homework must be turned in online through Courseworks in the format specified in the homework. Homeworks not adhering to these requirements will receive no credit.

Programming

Programming will be an integral part of this course, both for homeworks and projects. Students are strongly encouraged to use and contribute to the `GPy` project (available at: `https://github.com/SheffieldML/GPy`). `Matlab` and `R` are also allowed, though the student may find those choices to confer some disadvantage during later parts of the class. In other words, a small investment in a new computing platform early should pay off substantially by term's end (to say nothing of the career benefits of learning a more modern language). Because `python` and the `GPy` package may be new to some, the following resources are available to help the student make an informed choice about software:

- `https://github.com/cunni/gpkm/blob/master/getting_started.ipynb`
- `https://github.com/cunni/gpkm/blob/master/hw1.ipynb` (yes, this is meant to be the first homework; some small changes may occur by the time it is actually assigned)

Reading Material

Readings will be assigned alongside particular topics, and will often come from recent papers. You need not buy any of the books below, but a few helpful references are:

- Rasmussen, C.E. and Williams, C.K.I *Gaussian Processes for Machine Learning.* MIT Press, 2006. Available online at `http://gaussianprocess.org/gpml/`
- Schölkopf, B. and Smola, A.J. *Learning with Kernels.* MIT Press, 2001.
- Bogachev, V.I. *Gaussian Measures.* American Mathematical Society, 1998.

Approximate Lecture Outline (subject to change based on class)

Lecture	Content
1	Introduction to gaussian processes for machine learning
2	Kernels: commonly used, string, graph, Mallows, Kendall, mixture kernels, matrix manifold, etc.
3	Theory: existence, reproducing kernel Hilbert spaces, etc.
4	Speed and scaling part 1
5	Speed and scaling part 2
6	Approximate inference in gaussian process models
7	Dynamical systems with gaussian processes
8	Bayesian optimization and active learning
9	Probabilistic numerics: quadrature, ode solvers, etc.
10	(Project progress report)
11	Introduction to kernel methods
12	Kernel mean embeddings, and all that
13	Universal and characteristic kernels
14	Project final reports part 1
15	Project final reports part 2