Bayesian Machine Learning
ORIE 6741

Fall 2017
Tu/Th 11:40 am - 12:55 pm
Room: Rhodes Hall 571
`https://people.orie.cornell.edu/andrew/orie6741`

Syllabus

Instructor

Andrew Gordon Wilson
Assistant Professor
Rhodes Hall 235
Web: `https://people.orie.cornell.edu/andrew`
E-Mail: `andrew@cornell.edu`
Office Hours: Tuesday 4:00-5:00 pm, or by appointment

Course description:

To answer scientific questions, and reason about data, we must build models and perform inference within those models. But how should we approach model construction and inference to make the most successful predictions? How do we represent uncertainty and prior knowledge? How flexible should our models be? Should we use a single model, or multiple different models? Should we follow a different procedure depending on how much data are available?

In this course, we will approach these fundamental questions from a Bayesian perspective. From this perspective, we wish to faithfully incorporate all of our beliefs into a model, and represent uncertainty over these beliefs, using probability distributions. Typically, we believe the real world is in a sense *infinitely complex*: we will always be able to add flexibility to a model to gain better performance. If we are performing character recognition, for instance, we can always account for some additional writing styles for greater predictive success. We should therefore aim to maximize flexibility, so that we are capable of expressing any hypothesis we believe to be possible. For inference, we will not have a priori certainty that any one hypothesis has generated our observations. We therefore typically wish to weight an uncountably infinite space of hypotheses by their posterior probabilities. This *Bayesian model averaging* procedure has no risk of overfitting, no how matter how flexible our model. How we distribute our a priori support over these different hypotheses determines our *inductive biases*. In short, a model should distribute its support across as wide a range of hypotheses as possible, and have inductive biases which are aligned to particular applications.

This course aims to provide students with a strong grasp of the fundamental principles underlying Bayesian model construction and inference. We will go into particular depth on Gaussian process and deep learning models.

The course will be comprised of three units:

1. **Model Construction and Inference:** Parametric models, support, inductive biases, gradient descent, sum and product rules, graphical models, exact inference, approximate inference (Laplace approximation, variational methods, MCMC), model selection and hypothesis testing, Occam's razor, non-parametric models.
2. **Gaussian Processes:** From finite basis expansions to infinite bases, kernels, function space modelling, marginal likelihood, non-Gaussian likelihoods, Bayesian optimisation.
3. **Bayesian Deep Learning**: Feed-forward, convolutional, recurrent, and LSTM networks, dropout, GANs, variational autoencoders.

Depending on the available time, we may omit some of these topics. Most of the material will be derived on the chalkboard, with some supplemental slides.

Learning outcomes:

After taking this course, you should:

1. Be able to think about any problem from a Bayesian perspective.
2. Be able to create models with a high degree of flexibility and appropriate inductive biases.
3. Understand the interplay between model specification and inference, and be able to construct a successful inference algorithm for a given model.
4. Have familiarity with Gaussian process and deep learning models.

Course prerequisites:

I will assume solid knowledge of basic probability, linear algebra, and multivariable calculus. You should be comfortable with random variables, conditional probability and expectation, common probability distributions and their properties (binomial, geometric, exponential, Poisson, multivariate Gaussian), for example, as described in ORIE 3510. You should also be comfortable coding basic algorithms. We will mostly use Matlab or Python in the course; you will not be expected to already know these particular languages, but you should be able to pick them up independently at a basic level. Some prior exposure to machine learning and Bayesian statistics would be helpful, but not required. This course is designed to complement ORIE 4740, ORIE 6700, ORIE 6750, and machine learning courses in the Computer Science Department (CS 4780, 5780, 6784). Write me an e-mail if you are concerned about having the appropriate prerequisites. This course is suitable for graduate students at all levels. It will generally be presented at the beginning graduate level, and would be accessible to senior undergraduates.

Grading:

Five Assignments (20%)
One Midterm Exam (20%)
Readings (5%)
Final project (55%)

I will occasionally post reading materials, for which you are required to write a 1 page summary due at the beginning of class. These summaries comprise the 5% readings component of the grading scheme. As long as the readings are submitted on time (subject to a cumulative total of 5 grace days), and with a reasonable effort to understand the material, you will receive full credit.

The project will be graded as follows:

- Proposal: 10/100
- Midterm Report: 15/100
- Presentation: 15/100
- Peer Reviews: 20/100
- Final Report: 40/100

Our project website from last year provides additional information, though the details are subject to change: `https://people.orie.cornell.edu/andrew/orie6741/projinfoids.html`

Textbooks and readings:

There is no required textbook for the course. Some recommended textbooks:

- *Gaussian Processes for Machine Learning.* Rasmussen and Williams, MIT Press, 2006.
  Available free online: `http://www.gaussianprocess.org/gpml/`.
- *Information Theory, Inference, and Learning Algorithms.* MacKay, Cambridge University Press, 2003.
  Available free online: `http://www.inference.phy.cam.ac.uk/mackay/itila/book.html`.
- *Pattern Recognition and Machine Learning.* Bishop, Springer, 2006. You should ensure you have access to this book, e.g., through the library.

Assignments:

You are encouraged to discuss the problems with others, and to consult online sources. However, you must write your solutions independently and individually. Late homework will not be graded. If you cannot submit in person by 12 pm, then e-mail me a scan/clear picture of all pages of the homework by 12 pm, and then submit a physical copy as soon as possible for grading.

Exams:

One midterm exam. You are allowed to bring a double-sided sheet of notes.

Project:

In place of a final exam, there will be a final project near the end of the semester. This project involves formulating an original research proposal based on the content of the course, and then writing a report investigating this proposal. You are encouraged to consult with me or the TAs about the proposal. For the final report of the project, I would like an interactive document (using Matlab, Octave, or Python) which summarizes the problem and results, and has interactive demonstrations of the results.

Academic integrity:

You are expected to abide by the Cornell University Code of Academic Integrity. Any work submitted by you in this course for academic credit should be your own. The complete code is available at `http://cuinfo.cornell.edu/Academic/AIC.html`.

Disclaimer

This syllabus and grading scheme are subject to change.