Bayesian Machine Learning

Fall 2021
12-1:40pm, GCASL 361

Instructor

Andrew Gordon Wilson
60 Fifth Ave, Office 604
Web: `https://cims.nyu.edu/~andrewgw`
Office Hour: Tuesday 10:00-11:00 am

Teaching Assistants

Wesley Maddox
Office Hour: Friday 1:30-2:30 pm
Location: Bobst Library, LL138

Piazza

Please do not send e-mail. Instead, post questions to Piazza, or come to office hours. If you have a question, it is likely that others do too. Posting on Piazza is also a great way to start interacting with your classmates. Students are strongly encouraged to answer questions and initiate discussions on Piazza. General logistical requests, and questions about personal circumstances (e.g., “I need to connect by zoom, can you send me a link?”) should be made by direct message to instructors. Technical questions and posts related to the technical content of the course (such as a question about the readings or a pointer to an interesting article) should be public posts.
`piazza.com/nyu/fall2021/bayesianmachinelearningdsga3001013csciga3033087`
Please do not share this link with people not registered in the class.

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 processes. Note that this course is not intended to provide an introduction to deep learning, and there will likely be more material on Gaussian process models than neural networks. In terms of instruction the course will also largely be focused on conceptual considerations, rather than engineering or implementation.

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, MC Dropout, normalizing flows, deep kernel learning, loss surface geometry, subspace inference, function space perspectives.

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 modern Bayesian deep learning approaches.

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). 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.

You should carefully read the first three chapters of Bishop (2006), *Pattern Recognition and Machine Learning*, before starting the course. It is also highly recommended that you have taken prior machine learning courses.

This is a challenging and time-consuming course directed towards mature (independent) PhD students and researchers. You should only take it if you love the material, and have the time to devote to the course.

Grading:

Three Assignments (15%)
One Midterm Exam (20%)
Readings (10%)
Final project (55%)

I will also regularly post reading materials, for which you are required to write a 1 page summary due at the beginning of class. These summaries comprise the 10% readings component of the grading scheme. As long as the readings are submitted on time (subject to a cumulative total of 10 grace days), and with a reasonable effort to understand the material, you will receive full credit. Solutions will not be released for assignments, but they will be released for the midterm. This grading scheme is subject to change (in particular, the midterm may either not happen or change form, given the blended format of the course).

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. Available free online: `https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-` `pdf`

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.** All deliverables are submitted online through classes.

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.

The project will be graded as follows:

- Proposal: 10/100
- Midterm Report: 20/100
- Presentation: 15/100
- Peer Reviews: 20/100
- Final Report: 35/100

Our old project website provides additional information, though the details are subject to change: `https://cims.nyu.edu/~andrewgw/bml_project.html`

Academic integrity:

You are expected to abide by the NYU Code of Academic Integrity. Any work submitted by you in this course for academic credit should be your own.

Disclaimer

This syllabus and grading scheme are subject to change.