STAT547 – Bayesian Nonparametric Modelling and Inference
Fall 2018

Instructor: Trevor Campbell

Time & Place: Tuesday & Thursday, 14:30-16:00, ESB 4192

Course Website: `https://canvas.ubc.ca/courses/25155`

Description: How do we take full advantage of the nearly endless amount of data available from modern data sources, and make sure our models keep learning and refining what they've learned as we obtain more data? Bayesian nonparametric (BNP) models provide a solution: by employing an infinite-dimensional latent parameter (e.g. a function, measure, or sequence), BNP models continually grow and adapt to new observations, and have flexible, data-driven inference algorithms. This course covers state-of-the-art approaches to Bayesian nonparametrics, with a focus on modelling, inference algorithms, and foundational theory. The course will survey a number of models and algorithms for clustering, latent factor analysis, regression, topic modeling, and more, based on stochastic processes including the Dirichlet, Pitman-Yor Gaussian, Poisson, beta, and gamma. Different representations of each process (e.g. random measure, stick-breaking, marginalized predictive distributions) will be introduced along with their practical utility in learning algorithms. Posterior inference for each model via Markov chain Monte Carlo (MCMC) and variational methods will be covered. Grades will be based on a combination of homework assignments (involving light, guided theoretical exercises, coding exercises, and reading reflections) as well as a final course project.

Topics: The course will focus on Bayesian nonparametric modelling and inference algorithms, with some very light introduction to theoretical properties. A list of potential topics that will be covered follows.

- the basics of Bayes, parametric models, exponential families, conjugacy, inference algorithms
- Regression: the Gaussian process (GP), posterior inference
- Clustering & Mixture Models: the Dirichlet process (DP), representations, posterior inference
- Topic Modeling & HMMs: the hierarchical Dirichlet process (HDP), representations, posterior inference
- Binary Latent Factor Analysis: the beta process (BP), representations, posterior inference
- Integer Latent Factor Analysis: the gamma process (GaP), representations, posterior inference
- Theory: the Poisson process (PP) and (normalized) completely random measures

Prerequisites: There are no hard course prerequisites – consult the instructor if in doubt. Useful background to have:

- an introductory statistical modelling / probabilistic machine learning course (or equivalent)
- fluency in a general-purpose programming language (e.g. Python, C++, R).
- strong math background with sufficient linear algebra, calculus, and probability to read sophisticated research papers in machine learning and statistics conferences/journals
- Bayesian modelling and inference is not required & will be introduced, but familiarity coming in will be useful

Textbook/Materials: There is no required textbook. Course notes and papers will be made available on the course website.

Assessment: 60% homework assignments, 40% final project