

**University of California, Irvine
Statistics Distinguished Speaker Seminar**

MCMC Training of Bayesian Neural Networks

**Radford M. Neal
Professor Emeritus
University of Toronto**

11 a.m., Thursday, May 19, 2022

Bayesian training of neural network models avoids overfitting, provides proper quantification of uncertainty, and allows learning of high-level properties of the data, such as degree of smoothness, relevance of features, and additivity. For problems of small to moderate size, Bayesian training can be practically implemented using Markov chain Monte Carlo (MCMC) methods based on Hamiltonian dynamics, which computationally resembles batch gradient descent with momentum. But Bayesian inference for large networks, trained on large amounts of data, is computationally challenging. Exact MCMC methods require smaller stepsizes as the dimensionality of the parameter space increases. Simple methods for sampling hyperparameters also converge more slowly when each hyperparameter controls a large number of parameters. And batch updates looking at all training cases become infeasible when the training set is huge. I will discuss ways of alleviating these problems, such as by using good heuristics for relative stepsizes, using non-exact dynamical updates early in a run, clustering rejected trajectories, and computing trajectories with steps that use gradients for successive portions of the data. These methods are implemented in the latest release of my Software for Flexible Bayesian Modeling. This software can now also utilize GPUs for faster computation, and supports convolutional networks, allowing experiments with some standard deep learning tasks such as CIFAR-10.