Statistical Sciences
UNIVERSITY OF TORONTO

Feature allocations, paintboxes, and probability functions

Tamara Broderick, UC Berkeley

Clustering involves placing entities into mutually exclusive categories. We wish to relax the requirement of mutual exclusivity, allowing objects to belong simultaneously to multiple classes, a formulation that we refer to as "feature allocation." The first step is a theoretical one. In the case of clustering the class of probability distributions over exchangeable partitions of a dataset has been characterized (via exchangeable partition probability functions and the Kingman paintbox). These characterizations support an elegant nonparametric Bayesian framework for clustering in which the number of clusters is not assumed to be known a priori. We establish an analogous characterization for feature allocation; we define notions of "exchangeable feature probability functions" and "feature paintboxes" that lead to a Bayesian framework that does not require the number of features to be fixed a priori. The second step is a computational one. Rather than appealing to Markov chain Monte Carlo for Bayesian inference, we develop a method to transform Bayesian methods for feature allocation (and other latent structure problems) into optimization problems with objective functions analogous to K-means in the clustering setting. These yield approximations to Bayesian inference that are scalable to large inference problems.

Thursday February 27, 2014 at 3:30pm

Sidney Smith Hall, Room 1074

Refreshments will be served at 3:15pm