Bayesian Kernel Models: Theory and applications

Sayan Mukherjee

Department of Statistical Science
Institue for Genome Sciences & Policy
Department of Computer Science
Duke University

E-Mail: sayan@stat.duke.edu

Abstract: Kernel methods have been very popular in the machine learning literature in the last ten years, often in the context of Tikhonov regularization algorithms. I will introduce a coherent Bayesian kernel model based on an integral operator whose domain is a space of signed measures. Priors on the signed measures induce prior distributions on their image functions under the integral operator. I will identify general classes of measures whose images are dense in the reproducing kernel Hilbert space (RKHS) induced by the kernel. This gives a function-theoretic foundation for some nonparametric prior specifications commonly-used in Bayesian modeling, including Gaussian processes and Dirichlet processes, and suggests generalizations. A general framework for the construction of priors on signed measures using Levy processes is described.

An application of this model to high-dimensional gene expression data will illustrate how this Bayesian kernel model can be used to illustrate the "when, why, and how" the incorporation of unlabelled data, semi-supervised learning, helps in predictive regression models. This talk is based upon the following papers:

http://ftp.stat.duke.edu/WorkingPapers/06-18.html
http://www.imstat.org/sts/future_papers.html