



The University of Chicago
Department of Statistics

Seminar Series

MICHAEL JORDAN

Department of Electrical Engineering and Computer Science
Department of Statistics
University of California, Berkeley

“Hierarchical modeling with Dirichlet processes”

MONDAY, May 22, 2006 at 4:00 PM
133 Eckhart Hall, 5734 S. University Avenue
Refreshments following the seminar in Eckhart 110.

ABSTRACT

The Dirichlet process (DP) is a random measure that is useful in nonparametric Bayesian modeling, specifically in model-based clustering. In this presentation, I consider several extensions of the DP. Motivated by problems in bioinformatics and information retrieval, I first consider involving groups of clustering problems, in which it is desirable to share clusters among groups. While a simple Bayesian hierarchy in which multiple DP mixtures share a fixed base measure fail to yield the desired sharing, a more fully Bayesian notion of a “hierarchical Dirichlet process” (HDP) resolves the sharing problem. I discuss representations of the HDP based on extensions of stick-breaking and Chinese restaurant processes, and show how these representations yield MCMC algorithms for posterior inference. I also discuss alternatives to MCMC based on variational inference, showing that these methods can yield comparable accuracy at less computational cost. Finally, I discuss empirical Bayesian approaches to Dirichlet process mixtures which involve inference with respect to the base measure.

[Joint with Yee Whye Teh, Matt Beal, David Blei, Jon McAuliffe and Eric Xing].