The University of Chicago
Department of Statistics

Master's Thesis Presentation

RYAN WANG

Department of Statistics
The University of Chicago

Sparse Coding and an Application to Topic Modeling

THURSDAY, July 19, 2012, at 11:00 AM
110 Eckhart Hall, 5734 S. University Avenue

ABSTRACT

Sparse coding represents data as sparse linear combinations of basis vectors from a learned dictionary. Algorithms for sparse coding generally alternate between a coding step, which involves a regularized least squares problem, and a dictionary learning step, which involves a matrix factorization problem. In this paper we review some recent theoretical work on the performance of procedures used for solving these problems, such as the LASSO. We then present an application to topic modeling, a set of techniques for analyzing collections of text documents, and discuss some possible extensions that incorporate structured sparsity. Finally, we discuss experimental results from applying a sparse coding topic model to a collection of papers from the Neural Information Processing Systems (NIPS) conference.