The University of Chicago
Department of Statistics

MASTER'S THESIS PRESENTATION

VICTORIA YAO
Department of Statistics
The University of Chicago

Network Inference with Gene Expression Data Using Correlation and Regression-Based Methods

THURSDAY, May 12, 2011, at 1:30 PM
110 Eckhart Hall, 5734 S. University Avenue

ABSTRACT

The biological information that distinguishes different organisms is captured in their genetic material, but for this genetic information to be put in use, cells must undergo gene expression. Because gene expression is the key process in generating the basic building blocks of life, the processes that control gene expression are vital. One of the main ways genes are regulated is through transcription factors, which directly influence the process of transcription. Given gene expression data of both transcription factors and potential target genes (the result of gene regulation), we can attempt to reverse engineer the underlying gene regulatory networks. However, the noisy nature of the data and the complexity of the networks suggest endless possible methods and adjustments. We compared network inference using different correlation measures, including the Pearsons correlation, Spearmans correlation, and distance correlation. We also modeled gene expression as the response, where the covariates are the gene expression of transcription factors. We use LASSO regression to select for potential regulators in each network. Models were compared using the gold standards released by the DREAM5 Network Inference Challenge hosted by Columbia University and IBM.