Department of Statistics

DISSERTATION PROPOSAL

KUSHAL DEY

Department of Statistics
The University of Chicago

Pattern Detection and Visualization in Biological Data Using Grade of Membership Models

TUESDAY, November 1, 2016, at 2:00 PM
Jones 304, 5747 S. Ellis Avenue

ABSTRACT

Advances in genomic technologies has led to an explosion of biological data and sparked new challenges and opportunities for application of statistics and machine learning tools in the field. In this proposal talk, we focus on tools that can be used to model and visualize novel patterns in genomic and ecological data, that popular standard approaches often fail to represent effectively. We start with a discussion of a model-based clustering method, Latent Dirichlet Allocation developed for Natural Language Processing (Blei, Ng and Jordan 2003), that takes account of the count nature of data and allows each sample to have memberships in multiple clusters. We have shown the benefits of applying this model to detecting structural patterns in bulk and single- cell RNA sequencing (RNA-Seq) data, ancient DNA data, and to ecological data on birds abundance. We will discuss possible semi-supervised extensions of this model that supports data with partially known labels or biological information. We will also consider how this model can be extended to a paired factor analysis (PFA) approaches that can extract local graphical structure in the data. Finally, we will highlight the softwares we have developed for performing moving fit of the above methods and for visualizing the various structural patterns in RNA-Seq data, ecological data, and if time permits, DNA damage patterns in ancient DNA data.

For information about building access for persons with disabilities, please contact Laura Rigazzi at 773.702-0541 or send an email to lrigazzi@galton.uchicago.edu. If you wish to subscribe to our email list, please visit the following web site: https://lists.uchicago.edu/web/arc/statseminars.