Master’s Thesis Presentation

Daniel Li
Department of Statistics
The University of Chicago

“Simulation Study for Spatially Coherent Topic Modeling Utilizing Graph Structure”

April 26, 2024, at 10:00 AM
Jones 303, 5747 S. Ellis Avenue

Abstract

We examine a new technique for topic modeling that incorporates spatial information. Topic modeling is used for discovering the underlying themes in a collection of documents. While many applications of topic modeling focus on textual data, usage of topic modeling has been extended to analyzing cellular microenvironments. One recent method for microenvironment analysis is spatial latent Dirichlet allocation (SLDA), which uses a spatial prior in addition to the traditional LDA model. However, the spatial prior slows down computation. The method we will be examining, spatial probabilistic latent semantic indexing (SpLSI), is an SVD-based method that incorporates a graph regularization term to encourage neighboring environments to share topic compositions. Usage of the regularizer requires solving a convex clustering problem. Moreover, tuning the regularization necessitates a new cross validation scheme. We validate SpLSI through thorough simulation experiments and find improved performance over SLDA.