THE UNIVERSITY OF CHICAGO

Department of Statistics

DISSERTATION PRESENTATION AND DEFENSE

YOUNGSEOK KIM

Department of Statistics
The University of Chicago

"Bayesian Shrinkage Methods for High-dimensional Regression"

FRIDAY, April 30, 2020, at 3:00 PM
ZOOM Meeting

ABSTRACT

High dimensional data is prevalent in modern and contemporary science, and many statistics and machine learning problems can be framed as high dimensional regression — predicting response variables and selecting relevant features among candidates. One of the most commonly used Bayesian approaches is based on the i.i.d.~two component mixture prior, which comprises a point mass component at 0 ("spike") and a nonzero component ("slab"), on the regression coefficients. By computing the posterior probability that coefficients are zero, these Bayesian approaches allows us to measure the amount of shrinkage we need for individual regression coefficients and to identify relevant subsets of predictors. However, the above posterior inference can be done only approximately, and approximate inference procedures such as MCMC does not scale well to high dimensional and large data sets.

In this dissertation, we primarily focus on developing reliable Bayesian inferential tools that scale efficiently to dimensionality. Our first work proposes novel Variational Empirical Bayes (VEB) approaches to multiple linear regression based on a flexible scale mixture of normal distributions. The proposed approach (called Mr.ASH) is not only an approximate posterior inference procedure, but also a clever implementation of penalized regression where flexible EB replaces expensive cross validation. Our second work generalizes the two component mixture prior to the graph Laplacian prior, which accounts for graph structured sparsity. The general framework for Bayesian models, including sparse linear regression, change-point detection, clustering and more complex linear models with graph structured sparsity, will be presented. Our third work develops a fast algorithm for estimating mixture proportions, which serves as a central algorithmic routine in empirical Bayesian approaches to the normal means model and their applications such as linear regression and matrix factorization.

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