psd

DISSERTATION PRESENTATION AND DEFENSE

Empirical Bayes Matrix Factorization: Methods and Applications

WHEN
September 7, 2021
10:00 AM, CDT

WHERE
Via ZOOM
ZOOM information will be provided in the email announcement for this seminar.

Jason Willwerscheid, PhD candidate

Matrix factorization methods are commonly used to explore structure in multivariate data. When the structure can be expected to have a sparse representation, then a sparsity-inducing method will often be preferred. Empirical Bayes matrix factorization (EBMF), a recent approach that uses the observed data to estimate priors, can adaptively model sparsity and thus yield representations with interpretable components while also performing well on inferential tasks. Further, since fitting the EBMF model can be reduced to solving a series of empirical Bayes normal means (EBNM) subproblems, which can be relatively easily solved for a wide variety of prior families — sparse and otherwise —, the approach is very general.

The dissertation extends the reach of EBMF in several ways. The first chapter describes the R package **ebnm**, which I developed in order to provide a unified interface for efficiently solving the EBNM problem using a range of prior families. Existing packages are harnessed when practical; in other cases, solutions are implemented from scratch. The second chapter details my implementation of the EBMF algorithm, **flashier**, which was designed to handle much larger datasets and offer more flexibility than the original implementation of EBMF. In particular, I show that EBMF can yield insight into single-cell RNA sequencing data, outperforming more commonly used methods on tasks such as rare cell type detection in spite of the fact that the EBMF model is misspecified for count data. The final chapter considers data with an underlying tree-like structure, with particular attention to population genetics data. In addition to providing theory that helps to elucidate the kind of factorization that one should be looking for, I propose a tailored EBMF method that can successfully identify tree-like structure in both simulated and real datasets.

stat.uchicago.edu

THE UNIVERSITY OF **CHICAGO** | **DEPARTMENT OF STATISTICS**