Penalized Generative Variable Selection

Shuangge Ma

Department of Biostatistics, Yale School of Public Health

ABSTRACT

Deep networks are increasingly applied to a wide variety of data, including data with high-dimensional predictors. In such analysis, variable selection can be needed along with estimation/model building. Many of the existing deep network studies that incorporate variable selection have been limited to methodological and numerical developments. In this study, we consider modeling/estimation using the conditional Wasserstein Generative Adversarial networks. Group Lasso penalization is applied for variable selection, which can improve model estimation/prediction, interpretability, and stability. Significantly advancing from the existing literature, the analysis of censored survival data is also considered. We establish the convergence rate for variable selection while considering the approximation error, and obtain a more efficient distribution estimation. Simulations and the analysis of real experimental data demonstrate satisfactory practical utility of the proposed analysis.

Keywords: Deep neural network; Variable selection; Penalization; Consistency

Back to Sessions List

Sparse Representation Learning for Scalable Single-Cell RNA Sequencing Data Analysis

Kai Zhao**[1]**, Hon-Cheong So[2], **Zhixiang Lin[1]**

[1]*Department of Statistics, The Chinese University of Hong Kong, Hong Kong SAR, China*

[2]*School of Biomedical Sciences, The Chinese University of Hong Kong, Hong Kong SAR, China*

ABSTRACT

The rapid rise in the availability and scale of scRNA-seq data needs scalable methods for integrative analysis. Though many methods for data integration have been developed, few focus on understanding the heterogeneous effects of biological conditions across different cell populations in integrative analysis. Our proposed scalable approach, scParser, models the heterogeneous effects from biological conditions, which unveils the key mechanisms by which gene expression contributes to phenotypes. Notably, the extended scParser pinpoints biological processes in cell subpopulations that contribute to disease pathogenesis. scParser achieves favorable performance in cell clustering compared to state-of-the-art methods and has a broad and diverse applicability.

Keywords: Genomics; Integrative analysis; matrix factorization

Back to Sessions List

Bin Nan

Back to Sessions List