

On Sparse and Low-Rank Multivariate Models for Integrative Statistical Learning



Aditya Mishra, PhD
Computational Scientist
University of Texas MD Anderson Cancer Center

Modeling related responses using correlated predictors in high-dimensional settings is one of the most interesting problems in integrative statistical learning. Such problems are routinely required and formulated in various scientific investigations. For example, in genomics, understanding the role of transcription factors in regulating the gene expression of yeast cells observed at several time points; in gut microbiome data, understanding the association between host-associated features and microbial occurrence; and in cancer genomics, understanding the relationship between omics data including microbiome and metabolomics data. A multivariate analysis framework combined with a low-rank and sparse coefficient matrix is a promising way of learning the underlying association. However, learning such a decomposition is very challenging due to the simultaneous presence of orthogonality and sparsity constraints. Here, we introduce a divide-and-conquer strategy to infer such a coefficient matrix from data. In the divide step, we decompose the coefficient matrix into a sum of unit-rank matrices whose left and right singular vectors are sparse. Each unit-rank matrix is estimated using a sequential (greedy) approach in the conquer step. Depending on the outcome types, our approach assumes that conditionally independent observed responses follow a suitable exponential dispersion family. The proposed procedures are implemented in multiple user-friendly R packages. We have demonstrated the efficacy of the proposed procedures on multi-omics data from a colorectal cancer study, genomics data from a yeast cell cycle study, and microbiome-metabolites data from a Hepatocellular carcinoma study.

Monday, January 30, 2023 5:00-6:00 PM Eastern

W. Fred Mayes Telecommunication Center (230 Rosenau Hall)

Virtual using link and info below.

<https://unc.zoom.us/j/97439840799?pwd=V2tnbkI1aWwwR3ZoV01KNFFibkd2Zz09>

Meeting ID: 974 3984 0799

Passcode: 830399