



Mixed Graphical Models with Applications to Cancer Genomics

Genevera I. Allen*

Rice University & Baylor College of Medicine, Houston TX, USA - gallen@rice.edu

Eunho Yang & Pradeep Ravikumar

University of Texas, Austin TX, USA - e-mail address

Yulia Baker

Rice University, Houston TX, USA - e-mail address

Ying-Wooi Wan & Zhandong Liu

Baylor College of Medicine, Houston TX, USA - e-mail address

“Mixed Data”, comprising a large number of heterogeneous variables (e.g. count, binary, continuous, skewed continuous, among other data types), is prevalent in varied areas such as genomics, imaging genetics, national security, social networking, and Internet advertising. There have been limited efforts at statistically modeling such mixed data jointly, in part because of the lack of computationally amenable multivariate distributions that can capture direct dependencies between such mixed variables of different types. In this talk, we address this by introducing a novel class of heterogeneous so-called Block Directed Markov Random Fields (BDMRFs). Using the basic building block of node-conditional univariate exponential families from Yang et al. (2012), we introduce a class of mixed conditional random field distributions that are then chained according to a block-directed acyclic graph to form our class of Markov Networks. We characterize and study these models theoretically, propose penalized conditional likelihood estimators with statistical guarantees for recovering the underlying network structure, and study the estimators empirically. We conclude with an application to learning mixed genomic networks from next generation sequencing expression data and mutation data that discovers novel relationships between breast cancer biomarkers.

Keywords: Markov Networks; Mixed Big Data; Cancer Genomics; Graphical Models.