Weining Wang

Back to Sessions List

Covariance-Based Clustering and Biclustering via the Heterogeneous Block Covariance Model and Variants

Yunpeng Zhao, Xiang Li, Ning Hao, Qing Pan, and Mayson Zhang

Department of Statistics, Colorado State University

ABSTRACT

Clustering methods are traditionally designed to group data points in a (possibly high-dimensional) Euclidean space. We study a different but equally important task: clustering at the feature level, which arises naturally in gene expression analysis yet has received limited attention in statistics. In bioinformatics, it is often desirable to identify sets of highly correlated genes whose expression levels are jointly regulated and act synergistically to perform the same biological function. To address this problem, we introduce the heterogeneous block covariance model (HBCM), which characterizes community structure directly within the covariance matrix while accounting for heterogeneity in how features connect within communities. We develop a novel variational expectation–maximization algorithm for efficient estimation of group memberships. Theoretical analysis establishes the consistency of membership recovery, and simulation studies demonstrate the superior performance of HBCM compared to existing methods. We further extend HBCM to a joint clustering and biclustering framework that simultaneously partitions both features and samples. This generalization introduces new computational and theoretical challenges, for which we propose principled solutions. Applications to real gene expression data highlight the model's ability to uncover biologically meaningful clusters and biclusters.

Keywords: Variational EM algorithm; Covariance matrix; Biclustering; Gene expression data

Adaptive Block-Based Change-Point Detection for Sparse Spatially Clustered Data with Applications in Remote Sensing Imaging

Alan Moore, **Lynna Chu**, Zhengyuan Zhu

Department of Statistics, Iowa State University

ABSTRACT

We present a non-parametric change-point detection approach to detect potentially sparse changes in a time series of high-dimensional observations or non-Euclidean data objects. We target a change in distribution that occurs in a small, unknown subset of dimensions, where these dimensions may be correlated. Our work is motivated by a remote sensing application, where changes occur in small, spatially clustered regions over time. An adaptive block-based change-point detection framework is proposed that accounts for spatial dependencies across dimensions and leverages these dependencies to boost detection power and improve estimation accuracy. Through simulation studies, we demonstrate that our approach has superior performance in detecting sparse changes in datasets with spatial or local group structures. An application of the proposed method to detect activity, such as new construction, in remote sensing imagery of the Natanz Nuclear facility in Iran is presented to demonstrate the method's efficacy.

Keywords: Change-point; Non-parametric; Graph-based tests; Spatial dependence; Satellite images

Graph Release with Assured Node Differential Privacy

Suqing Liu, Xuan Bi, **Tianxi Li**[1]

University of Chicago and University of Minnesota, Twin Cities

ABSTRACT

Differential privacy is a well-established framework for safeguarding sensitive information in data. While extensively applied across various domains, its application to network data --- particularly at the node level --- remains underexplored. Existing methods for node-level privacy either focus exclusively on query-based approaches, which restrict output to pre-specified network statistics, or fail to preserve key structural properties of the network. In this work, we propose GRAND (Graph Release with Assured Node Differential privacy), which is, to the best of our knowledge, the first network release mechanism that releases entire networks while ensuring node-level differential privacy and preserving structural properties. Under a broad class of latent space models, we show that the released network asymptotically follows the same distribution as the original network. The effectiveness of the approach is evaluated through extensive experiments on both synthetic and real-world datasets.

Keywords: Data Privacy; Privacy-Preserving Networks; Data Sharing; Network Release

Back to Sessions List