



THE HONG KONG UNIVERSITY OF SCIENCE & TECHNOLOGY

Department of Mathematics

SEMINAR ON STATISTICS

Bayesian Pyramids: Identifiable Multilayer Discrete Latent Structure Models for Discrete Data

By

Prof. Yuqi GU
Columbia University

Abstract

High dimensional categorical data are routinely collected in biomedical and social sciences. It is of great importance to build interpretable parsimonious models that perform dimension reduction and uncover meaningful latent structures from such discrete data. Identifiability is a fundamental requirement for valid modeling and inference in such scenarios, yet is challenging to address when there are complex latent structures. In this article, we propose a class of identifiable multilayer (potentially deep) discrete latent structure models for discrete data, termed Bayesian pyramids. We establish the identifiability of Bayesian pyramids by developing novel transparent conditions on the pyramid-shaped deep latent directed graph. The proposed identifiability conditions can ensure Bayesian posterior consistency under suitable priors. As an illustration, we consider the two-latent-layer model and propose a Bayesian shrinkage estimation approach. Simulation results for this model corroborate the identifiability and estimability of model parameters. Applications of the methodology to DNA nucleotide sequence data uncover useful discrete latent features that are highly predictive of sequence types. The proposed framework provides a recipe for interpretable unsupervised learning of discrete data, and can be a useful alternative to popular machine learning methods.

Date : 25 March 2022 (Friday)

Time : 10:00am

Zoom Meeting : <https://hkust.zoom.us/j/6827297694> (Passcode: 7436)

All are Welcome!