INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO

Departamento Académico de Estadística
Río Hondo # 1, Col. Progreso Tizapán,
Ciudad de México, C.P. 01080, México

Seminario Aleatorio

Sesión 399

Multi-study Factor Regression Models for Large Complex Data with Applications to Nutritional Epidemiology and Cancer Genomics

Alejandra Ávalos Pacheco
https://sites.google.com/view/aleavalos

Abstract

Data-integration of multiple studies can be key to understand and gain knowledge in statistical research. However, such data present both biological and artifactual sources of variation, also known as covariate effects. Covariate effects can be complex, leading to systematic biases. In this talk I will present novel sparse latent factor regression (FR) and multi-study factor regression (MSFR) models to integrate such heterogeneous data. The FR model provides a tool for data exploration via dimensionality reduction and sparse low-rank covariance estimation while correcting for a range of covariate effects, such as batch effects. MSFR are extensions of FR that enable us to jointly obtain a covariance structure that models the group-specific covariances in addition to the common component. I will discuss the use of several sparse priors (local and non-local) to learn the dimension of the latent factors. Our approach provides a flexible methodology for sparse factor regression which is not limited to data with covariate effects. I will present several examples, with a focus on bioinformatics applications. We show the usefulness of our methods in two main tasks: (1) to give a visual representation of the latent factors of the data, i.e. an unsupervised dimension reduction task and (2) to provide a (i) supervised survival analysis, using the factors obtained in our method as predictions for the cancer genomic data; and (ii) dietary pattern analysis, associating each factor with a measure of overall diet quality related to cardio-metabolic disease risk for a Hispanic community health nutritional-data study.Our results show an increase in the accuracy of the dimensionality reduction, with non-local priors substantially improving the reconstruction of factor cardinality. The results of our analyses illustrate how failing to properly account for covariate effects can result in unreliable inference.

Paper: https://projecteuclid.org/journals/bayesian-analysis/volume-17/issue-1/Heterogeneous-Large-Datasets-Integration-Using-Bayesian-Factor-Regression/10.1214/20-BA1240.full

Viernes 03 de febrero de 2023,
13:00 horas de CDMX,

https://itam.zoom.us/j/91249364522?pwd=Rk0xMFY3bEE5bFFsSWZRc1h1ZTljQT09
ID de reunión: 912 4936 4522
Código de acceso: 648250

El Seminario Aleatorio del Departamento de Estadística del ITAM está destinado tanto a profesores como a estudiantes, por lo que se agradece a los profesores que colaboren invitando a sus alumnos a estas sesiones.