ALPEN-ADRIA UNIVERSITÄT KLAGENFURT

INVITATION TO THE DOCTORAL SEMINAR

Scalable Integrative Factor Models: From Nutritional Epidemiology to Cancer Genomics

Alejandra Avalos Pacheco | Johannes Kepler Universität Linz

Chaired by Michaela Hitz | Department of Statistics

Wednesday, 28 October 2026 | 10:00–11:00

HS 11

ABSTRACT

The integration of data across studies and populations offers important opportunities for understanding complex biological systems, but heterogeneity, covariate effects, and other systematic sources of variation can obscure the underlying structure and lead to biased or unreliable inference.

In this talk, I will present a class of sparse latent factor models for the analysis and integration of heterogeneous high-dimensional data, together with scalable computational methods for their estimation. I will first introduce factor regression (FR) and multi-study factor regression (MSFR) models, which provide flexible approaches to dimensionality reduction and sparse low-rank covariance estimation while accounting for covariate effects. MSFR extends this framework by jointly modelling latent covariance structures that are shared across studies together with study-specific sources of variation.

I will discuss the use of local and non-local sparse priors for learning both interpretable latent factors and their dimensionality, and illustrate how these

models can be used for unsupervised data exploration as well as for downstream supervised analyses.

I will then present a recent extension, the Integrative Rotationally Invariant Sparse Factor Model (IRIS-FM), which is designed to estimate sparse and interpretable common and study-specific factors while adjusting for confounding effects and avoiding restrictive identifiability constraints and post-processing rotations. Its application to human and mouse hepatocellular carcinoma gene-expression data allows us to characterize shared and species-specific genomic structure, assess which mouse models best represent different human patient subpopulations, and identify co-regulatory mechanisms that may or may not translate across species.

Applications in nutritional epidemiology and cancer genomics will illustrate how these models can recover meaningful latent structure, identify dietary patterns associated with cardiometabolic risk, support survival prediction, and improve the assessment of biological similarities across studies and species.

We hope to see you there!