DaSSWeb
Data Science and Statistics Webinar

28TH JANUARY | 2.30 PM (GMT) | JOIN US HERE

MODEL-BASED CO-CLUSTERING: HIGH DIMENSION AND ESTIMATION CHALLENGES

Model-based co-clustering can be seen as a particularly important extension of model-based clustering. It allows for a significant reduction of both the number of rows (individuals) and columns (variables) of a data set in a parsimonious manner, and also allows interpretability of the resulting reduced data set since the meaning of the initial individuals and features is preserved. Moreover, it benefits from the rich statistical theory for both estimation and model selection. Many works have produced new advances on this topic in recent years, and we offer a general update of the related literature. It is the opportunity to advocate two main messages, supported by specific research material: (1) co-clustering requires further research to fix some well-identified estimation issues, and (2) co-clustering is one of the most promising approaches for clustering in the (very) high-dimensional setting, which corresponds to the global trend in modern data sets.
It is a joint work with Christine Keribin (U. Paris-Saclay) and Julien Jacques (U. Lyon 2).

SPEAKER

CHRISTOPHE BIERNACKI | LAB. MATHEMATICS, UNIVERSITÉ LILLE 1 & INRIA LILLE NORD-EUROPE FRANCE

Christophe Biernacki is a senior researcher at Inria, which is the National Institute for Research in Digital Science and Technology in France, where he acts also as a deputy scientific director in charge of applied mathematics at the national level. His research topics concern model-based clustering for complex data cases, including for instance mixed data or HD data. He has also a look at some fundamental questions in clustering as making a bridge between model-based and deep-based clustering, or also addressing frugality for clusterizing tall data sets.

FEP