UNIVERSITÀ
DI PAVIA

Ph.D. Program in Electronics, Computer Science and Electrical Engineering

SEMINAR

Model-based clustering of categorical data based on the Hamming distance

LUCIA PACI
UNIVERSITA' CATTOLICA DEL SACRO CUORE, MILANO

13th May 2024/12.30
Common Room, Department of Economics and Management, Via San Felice 5

Abstract:
A model-based approach is developed for clustering categorical data with no natural ordering. The proposed method exploits the Hamming distance to define a family of probability mass functions to describe vectors of categorical data. The elements of this family are then considered as kernels of a finite mixture model with an unknown number of components. Conjugate Bayesian inference has been derived for the parameters of the Hamming distribution model. The mixture is framed in a Bayesian nonparametric setting, and a transdimensional blocked Gibbs sampler is developed to provide full Bayesian inference on the number of clusters, their structure and the group-specific parameters, facilitating the computation with respect to customary reversible jump algorithms. The proposed model encompasses a parsimonious latent class model as a special case when the number of components is fixed. Model performances are assessed via a simulation study and reference datasets, showing improvements in clustering recovery over existing approaches. This is joint work with Raffaele Argiento and Edoardo Filippi-Mazzola.

Bio:
Lucia Paci is Associate Professor of Statistics. Main research topics: Bayesian inference of spatio-temporal modeling, Graphical models, Ecological and Environmental applications.

Organizer	**Ph.D. Coordinator**
Prof. Paolo Giudici	Prof.ssa Ilaria Cristiani