

Hybrid approach for the analysis of complex data structures

§ Name of instructor

Alfonso Iodice D'Enza, University of Naples Federico II, Italy

§ Short description

In real data structures, complexity comes from different aspects: attributes of different nature (continuous and categorical), blocks of attributes that refer to a same latent concept, groups of homogeneous observations. The statistical learning of such data structures require the application of suitable combinations of existing methods to take into account the complexity and/or to reduce the complexity: the approaches in question are referred to as hybrid. Examples of hybrid approaches are principal component and partial least squares regression: the starting attributes are used to define synthetic features that are then used as predictors to model the response of interest. In complex surveys, where sets of items refer to a same latent concept, hybrid approaches can be used to define a single synthetic response out of a set of items, and then to regress it on other items/attributes at hand.

§ Schedule

1. Model complexity and the bias variance trade-off
2. Reducing model complexity
3. Modeling meta-responses
4. Hybrid unsupervised approaches

§ Introductory background

Gareth, J., Daniela, W., Trevor, H., & Robert, T. (2013). *An introduction to statistical learning: with applications in R*. Springer.

Iannario, M., D'Enza, A. I., & Romano, R. (2022). A hybrid approach for the analysis of complex categorical data structures: assessment of latent distance learning perception in higher education. *Computational Statistics*, 1-19.

§ Facilities Required

- Software: R, open source, with the R suites tidyverse and tidymodels
- Course Material. All course materials, including the data and R scripts for the examples, will be made available for course participants.