

N-CovSel, a new strategy for feature selection in N-way data

Alessandra Biancolillo¹, Federico Marini², Jean-Michel Roger³

¹*University of L'Aquila, Via Vetoio, 67100, Coppito, L'Aquila, Italy*

²*University of Rome "La Sapienza", Piazzale Aldo Moro 5, 00185, Rome, Italy*

³*ITAP, Inrae, Montpellier SupAgro, University of Montpellier, Montpellier, France*

e-mail: jean-michel.roger@inrae.fr

In data analysis, how to select meaningful variables is a hot and wide-debated topic and several variable selection (or feature reduction) approaches have been proposed into the literature.

Covariance Selection (CovSel) is an existing method which is conceived to select variables in regression and discrimination contexts, and it assesses the features' relevancy based on their covariance with the response(s). Most of variable selection methods refer to contexts in which data are collected in matrices. How to assess the relevancy of variables in a multi-way context has not been extensively discussed yet.

The present contribution, named N-CovSel, proposes to extend the CovSel principle to the N-Way structures, by selecting features in place of variables. Three main questions are addressed to achieve this: (i) How to define a feature in a N-Way array (Figure 1); (ii) How to define the covariance between a feature and a response Y; (iii) How to deflate a N-Way array with regard to a selected feature.

The complete algorithm of N-CovSel will be presented and its theoretical properties discussed. Two applications on 3-way real data will be presented, illustrating that the proposed method can be differently used, depending on the final purpose of the analysis.