

## **Supervised-Component-Based Generalised Linear Regression with Multiple Explanatory Blocks: THEME-SCGLR**

**Authors :** Bry X., Trottier C., Mortier F., Cornu G., Verron T.

**Abstract :** We address component-based regularization of a Multivariate Generalized Linear Model (MGLM). A set of random responses  $Y$  is assumed to depend, through a GLM, on a set  $X$  of explanatory variables, as well as on a set  $T$  of additional covariates.  $X$  is partitioned into  $R$  conceptually homogeneous blocks  $X_1, \dots, X_R$ , viewed as explanatory themes. Variables in each  $X_r$  are assumed many and redundant. Thus, Generalised Linear Regression (GLR) demands regularization with respect to each  $X_r$ . By contrast, variables in  $T$  are assumed selected so as to demand no regularization. Regularization is performed searching each  $X_r$  for an appropriate number of orthogonal components that both contribute to model  $Y$  and capture relevant structural information in  $X_r$ . We propose a very general criterion to measure structural relevance (SR) of a component in a block, and show how to take SR into account within a Fisher-scoring-type algorithm in order to estimate the model. We show how to deal with mixed-type explanatory variables. The method, named THEME-SCGLR, is tested on simulated data.

**Keywords :** Component-Model, Fisher Scoring Algorithm, GLM, PLS Regression, SCGLR, SEER, THEME

**Conference Title :** ICCSDA 2015 : International Conference on Computational Statistics and Data Analysis

**Conference Location :** Istanbul, Türkiye

**Conference Dates :** January 26-27, 2015