

Selection of fixed effects in high dimensional linear mixed models using a multicycle ECM algorithm

Magali San Cristobal

Laboratoire de Génétique Ceullaire (LCG), INRA, Toulouse, France

Linear mixed models are especially useful when observations are grouped. In a high dimensional setting however, selecting the fixed effect coefficients in these models is mandatory as classical tools are not performing well. By considering the random effects as missing values in the linear mixed model framework, a ℓ_1 -penalization on the fixed effects coefficients of the resulting log-likelihood is proposed. The optimization problem is solved via a multicycle Expectation Conditional Maximization (ECM) algorithm which allows for the number of parameters p to be larger than the total number of observations n and does not require the inversion of the sample $n \times n$ covariance matrix. The proposed algorithm can be combined with any variable selection method developed for linear models. A variant of the proposed approach replaces the ℓ_1 -penalization with a multiple testing procedure for the variable selection aspect and is shown to greatly improve the False Discovery Rate. Both methods are implemented in the MMS R-package, and are shown to give very satisfying results in a high-dimensional simulated setting.