

DSS Statistics Seminar

March 19, 2021, 12:00

<https://uniroma1.zoom.us/j/86881977368?pwd=SWRFcVFjMDZTa0lXZk05TE1zNm5adz09>

Passcode: 432940

LASSO-penalized clusterwise
linear regression modeling with
Local Least Angle Regressions

R. Di Mari¹, joint work with R. Rocci², A.S. Gattone³

¹University of Catania, ²La Sapienza University, ³D'Annunzio University

In clusterwise regression analysis, the goal is to predict a response variable based on a set of explanatory variables, where each predictor has different contributions to the response depending on the cluster. The number of candidates is typically large: whereas some of these variables might be useful, some others might contribute very little to the prediction. A well known method to perform variable selection is the LASSO, where the penalty is calibrated by minimizing the Bayesian Information Criterion (BIC). However, available approaches to the computation of LASSO-penalized estimators are time consuming and/or require approximate schemes making the tuning of the penalty cumbersome. In order to ease such computation, we introduce an expectation maximization algorithm with closed-form updates. This is based on an iterative scheme where the component specific lasso regression coefficients are computed according to a coordinate descent soft-thresholding update, and the LARS algorithm is used to derive the full path of component specific LASSO solutions for model selection. We show the advantage of this approach in terms of model selection and computation time reduction by means of a simulation study, and illustrate it with an application to a Major League Baseball salary data.



SAPIENZA
UNIVERSITÀ DI ROMA