Regularized Nonhomogeneous Regression for Predictor Selection in Ensemble Post-Procession

Jakob W. Messner (1), Georg J. Mayr (2), and Achim Zeileis (2)

(1) Technical University of Denmark, Kgs. Lyngby, Denmark (jwmm@elektro.dtu.dk), (2) University of Innsbruck, Innsbruck, Austria

Non-homogeneous regression is a popular method to post-process ensemble forecasts. Usually only ensemble forecasts of the predictand variable are used as input but other potential predictor variables such as current observations, ensemble forecasts of other variables or for other locations, transformations of those, etc. are disregarded. Although it is straightforward to incorporate other predictor variables in nonhomogeneous regression these have to be selected carefully to avoid overfitting.

In this contribution we present two approaches to automatically select the most relevant predictor variables in nonhomogeneous regression. Gradient boosting and lasso regularization are alternative approaches to estimate the regression coefficients and both avoid overfitting by regularizing the coefficients. The most important predictor variables are selected automatically by restricting coefficients of less important variables to zero.

A case study with ensemble forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF) shows that both approaches effectively select important input variables and can clearly improve the forecasts. The performance differences between the two approaches are small but lasso regularization is computationally more efficient while gradient boosting is more flexible.