Using dimension reduction for quantifying and estimating predictability and explainability in regression analysis

Sebastian Fuchs*

Universität Salzburg
Salzburg, Austria

Abstract

Two omnipresent concepts in the context of regression are *predictability* and *explainability*. Predictability involves quantifying the extent of functional (or monotone regression) dependence of the endogenous variable Y on a set of $d \geq 1$ potential exogenous variables $X_1, \ldots, X_d$, mostly with the aim to select the most relevant explanatory variables. Explainability refers to the quantification of the fraction of explained variance and quantitatively assesses the appropriateness of a chosen statistical model.

Choosing the 'simple measure of conditional dependence' T recently introduced by Azadkia & Chatterjee (2021) as a measure for predictability and the multivariate extension of copula correlation as a measure for explainability allows to work with a dimension reduction principle. We identify and investigate the dependence structure underlying this dimension reduction principle, provide a strongly consistent estimator for the aforementioned quantities, and demonstrate its broad applicability.

*E–mail Sebastian Fuchs: `sebastian.fuchs@plus.ac.at`