European Network for Business and Industrial Statistics

Contribution ID: **94**   Type: **not specified**

Combining Statistical Rigor with Machine Learning: Split-Based Forward Selection method to choose the most important variables

Real world datasets frequently include not only vast numbers of observations but also high dimensional feature spaces. Exhaustively gathering and examining every variable to uncover meaningful insights can be time consuming, costly, or even infeasible. In order to build up robust, reliable and efficient regression models, feature selection techniques have therefore become inevitable. Yet many established methods consider only a single partition of the training data, risking biased or sub optimal variable choices. This work introduces a modified forward selection strategy for feature selection executed over multiple data splits, accounting for inter variable relationships and structural variation within the data to highlight the most influential variables.

Special/ Invited session

Classification

Both methodology and application

Keywords

Machine Learning, Feature Selection, Multicollinearity

Primary author: BIRIS, Panagiotis (University of Patras)

Co-author: Dr ECONOMOU, Polychronis

Presenter: BIRIS, Panagiotis (University of Patras)

Track Classification: Machine Learning