Contribution ID: **79** Type: **not specified**

Self-Validated Ensemble Models (SVEM) –Machine Learning for Small Data Typical of Industrial Designed Experiments

Tuesday, 12 September 2023 17:00 (20 minutes)

Self-Validating Ensemble Modeling (S-VEM) is an exciting, new approach that combines machine learning model ensembling methods to Design of Experiments (DOE) and has many applications in manufacturing and chemical processes. In most applications, practitioners avoid machine learning methods with designed experiments because often one cannot afford to hold out runs for a validation set without fundamentally changing the aliasing structure of the design. We present a technique that fractionally allocates rows to training and validation sets that makes machine learning model selection possible for the small datasets typical in DoE applications. The approach with S-VEM is similar to Random Forests ™ except that instead of averaging a set of resampling-based bootstrap decision tree models, one averages fractional-random-weight bootstrap linear models whose effects have been chosen using forward selection or the Lasso. In this way, we are able to retain the interpretability of response surface models, while being able to obtain greater accuracy as well as fit models that have more parameters than observations. Although our investigations have only applied the S-VEM model averaging technique to linear least squares models, the algorithm is quite general and could be applied to generalized linear models, as well as other machine learning methods like neural networks or decision trees. We will present simulation results comparing independent test set accuracy of S-VEM to more traditional approaches to modeling DoE data and illustrate the method with case studies.

Keywords

DOE, Machine Learning

Classification

Both methodology and application

Primary author: Dr GOTWALT, Christopher (JMP Statistical Discovery LLC)

Presenter: Dr GOTWALT, Christopher (JMP Statistical Discovery LLC)

Session Classification: CONTRIBUTED Special Session: Design of Experiments

Track Classification: Machine learning