

Repro Samples Method for Addressing Irregular Inference Problems and for Unraveling Machine Learning Blackboxes

Min-ge Xie
Rutgers University, USA

Abstract

Rapid data science developments and the desire to have interpretable AI require us to have innovative frameworks to tackle frequently seen, but highly non-trivial “irregular inference problems,” e.g., those involving discrete or non-numerical parameters and those involving non-numerical data, etc. This talk presents an effective and wide-reaching framework, called repro samples method, to conduct statistical inference for the irregular problems and more. We develop both theories to support our development and provide effective computing algorithms for problems in which explicit solutions are not available. The method is likelihood-free and is particularly effective for irregular inference problems. For commonly encountered irregular inference problems that involve discrete or nonnumerical parameters, we propose a three-step procedure to make inferences for all parameters and develop a unique matching scheme that turns the disadvantage of lacking theoretical tools to handle discrete/nonnumerical parameters into an advantage of improving computational efficiency. The effectiveness of the proposed method is illustrated through case studies by solving two highly nontrivial problems in statistics: a) how to quantify the uncertainty in the estimation of the unknown number of components and make inference for the associated parameters in a Gaussian mixture; b) how to quantify the uncertainty in model estimation and construct confidence sets for the unknown true model, the regression coefficients, or both true model and coefficients jointly in high dimensional regression models. The method also has extensions to complex machine learning models, e.g., (ensemble) tree models, neural networks, graphical models, etc. It provides a new toolset to develop interpretable AI and to help address the blackbox issues in complex machine learning models.