Repro Sampling Method for Joint Inference of Model Selection and Regression Coefficients in High Dimensional Linear Models

Peng Wang

University of Cincinnati
E-mail: wangp9@ucmail.uc.edu

Abstract: This paper proposes a new and effective simulation-based approach, called Repro Sampling method, to conduct statistical inference in high dimensional linear models. The Repro method creates and studies the performance of artificial samples (referred to as Repro samples) that are generated by mimicking the sampling mechanism that generated the true observed sample. By doing so, this method provides a new way to quantify model and parameter uncertainty and provide confidence sets with guaranteed coverage rates on a wide range of problems. A general theoretical framework and an effective Monte-Carlo algorithm, with supporting theories, are developed for high dimensional linear models. This method is used to joint create confidence sets of selected models and model coefficients, with both exact and asymptotic inferences are included. It also provides theoretical development to support the computational efficiency. Furthermore, this development allows us to handle inference problems involving covariates that are perfectly correlated. A new and intuitive graphical tool to present uncertainties in model selection and regression parameter estimation is also developed. We provide numerical studies to demonstrate the utility of the proposed method in a range of problems. Numerical comparisons suggest that the method is far better (in terms of improved coverage rates and significantly reduced sizes of confidence sets) than the approaches that are currently used in the literature. The development provides a simple and effective solution for the difficult post-selection inference problems.