Abstract Archives

"But what about interactions, are any of those significant?" Resolving the Collaborative Nightmares of Covariate Interactions through Regularization

Abstract Body: Sifting through all possible interactions in modeling applications is a dangerous statistical endeavor. All too often, one or more of the many interactions is found to "significantly" improve the fit, and we burden ourselves with trying to interpret an opaque model with interactions that do not make clinical sense. With this in mind, we explore and illustrate the concept of ranked sparsity, a phenomenon that often occurs naturally in the presence of derived variables such as interactions. In particular, ranked sparsity arises in modeling applications when an expected disparity exists in the quality of information between different feature sets. Its presence can cause traditional and modern model selection methods to fail because such procedures commonly presume that each potential parameter is equally worthy of entering into the final model – we call this presumption "covariate equipoise". However, when all possible interactions are considered as candidate predictors, the premise of covariate equipoise will often produce over-specified and convoluted models. The sheer number of additional candidate variables grossly inflates the number of false discoveries in the interactions, resulting in unnecessarily complex and difficult-to-interpret models with many (truly spurious) interactions. We suggest a modeling strategy that requires a stronger level of evidence in order to allow certain variables (e.g. interactions) to be selected in the final model. This ranked sparsity paradigm can be implemented with the sparsity-ranked lasso (SRL). We compare the performance of SRL relative to competing methods for selecting interactions in a series of simulation studies, showing that the SRL is fast, accurate, and produces more transparent models (with fewer false interactions) than other state-of-the-art methods. We illustrate its utility in an application to predict the survival of lung cancer patients using a set of gene expression measurements and clinical covariates, searching in particular for gene-environment interactions, which are very difficult to find in practice.

Meeting: 2020 Virtual International Biometric Conference
Methodological Topic: Regression modelling
Application Area: Medical research

Authors

- Peterson, Ryan (University of Colorado School of Public Health , Denver , Colorado , United States)
- Cavanaugh, Joseph (University of Iowa College of Public Health , Iowa City , Iowa , United States)

SESSION INFORMATION

Poster Session - Group 1

Poster Session - Group 1