

Title: Regularization, predictions, and sample selection model

Emmanuel Ogundimu

General description

Sample selection arises when the outcome of interest is partially observed in a study. The data are often realized through a two-stage process, a first-level binary mechanism which determines whether an observation will be available for the second level. Consider, as an example, where a loan application is made to a bank. The bank uses the loan applicant attributes to grant or reject the loan request. If the request is accepted, then the bank will observe the loan performance over time. These involve a two-stage process, the credit granting process (accept or reject) and loan performance process (default or non-default). A model developed using the accept only applicants from the credit granting process may be a non-random sample from the target population and can lead to selection bias.

Objectives

This project will involve the development of regularization techniques such as LASSO (**least absolute shrinkage and selection operator**) for variable selection and optimization of predictive accuracy of models that are developed in this framework. Some potential directions could be:

- Development of estimators based on LASSO and Adaptive LASSO using the least square approximation method
- Development and study of a new estimator for LASSO which is based on the approximation of L_1 -norm and post selection inference
- Extension to various penalty functions, e.g. MCP (minimax concave penalty), SCAD (smoothly clipped absolute deviation).
- Robustification via student-t error distribution
- Missing data imputation and interesting data application in reject inference problems.
- Extension of validation methods to regularized sample selection model.

Prerequisite skills: Calculus, Newton-Raphson method, Generalized Linear Models, Programming in R,

References

- (1) Greene, W. H. (1998). Sample selection in credit scoring models. *Jpn. World Econ* 10, 299--316.
- (2) Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Series B* 58(1), 267–288.
- (3) Zou, H. (2006). The Adaptive Lasso and Its Oracle Properties. *J. Am. Stat. Assoc.* 101(476), 1418–1429.