

Robust logistic regression with sparse predictor variables

Authors: Ana Bianco¹, Graciela Boente¹ and Gonzalo Chebi¹

¹Universidad de Buenos Aires and CONICET, Argentina

Abstract:

Nowadays, dealing with high-dimensional data is a recurrent problem that cuts across modern statistics. One main feature of high dimensional data is that the dimension p , that is, the number of covariates, is high, while the sample size n is relatively small. In this circumstance, the bet on sparsity principle suggests to proceed under the assumption that most of the effects are not significant. Sparse covariates are frequent in the classification problem and in this situation the task of variable selection may be also of interest.

We focus on the logistic regression model and our aim is to address robust and sparse estimators of the regression parameter in order to perform estimation and variable selection at the same time.

For this purpose, we introduce a family of penalized M-type estimators for the logistic regression parameter that are stable against atypical data. We explore different penalizations functions and we introduce the so-called sign penalization. This new penalty has the advantage that it does not shrink the estimated coefficients to 0 and that it depends on only one parameter.

We will discuss some the variable selection capability of the proposal as well as its asymptotic behaviour. Through a numerical study, we compare the finite sample performance of the proposal with different penalized estimators either robust and classical under different scenarios.