

Variable selection via SCAD-penalized quantile regression for high-dimensional count data

ABSTRACT

This article introduces a quantile penalized regression technique for variable selection and estimation of conditional quantiles of counts in sparse high-dimensional models. The direct estimation and variable selection of the quantile regression is not feasible due to the discreteness of the count data and non-differentiability of the objective function, therefore, some smoothness must be artificially imposed on the problem. To achieve the necessary smoothness, we use the Jittering process by adding a uniformly distributed noise to the response count variable. The proposed method is compared with the existing penalized regression methods in terms of prediction accuracy and variable selection. We compare the proposed approach in zero-inflated count data regression models and in the presence of outliers. The performance and implementation of the proposed method are illustrated by detailed simulation studies and real data applications.

Keyword: Count data; High dimensional; Jittering; Quantile regression; Variable selection; Zero-inflated