Parsimonious additive logistic models

Marta Avalos

Parsimonious additive logistic models

Logistic regression is a standard tool in statistics for binary classification. The logistic model relates the logarithm of the odds-ratio to the predictors via a linear regression model. A generalization is the additive logistic model, which replaces each linear term by an unspecified smooth function, allowing for more flexibility while preserving interpretability. Another variant is penalized logistic regression, which shrinks coefficients to improve the accuracy of prediction. Ridge regression (L2-penalization) and lasso (L1-penalization) are the main penalization procedures. An attractive property of the later is that it performs parameter estimation and variable selection simultaneously. New theoretical results, efficient algorithms, and available software play a major role in the recent popularization of lasso. In this study, L1-penalization is adapted to additive logistic regression fitted by smoothing splines. Coefficients associated to predictors. Coefficients associated to predictors with little effect on the response may be shrunk (some of them to zero). This approach gives parsimonious models, removes irrelevant variables, and identifies non linear trends. The estimates are computed via the usual Newton-Raphson update, combined with the lars-lasso algorithm, to resolve the penalization problem, and the backfitting algorithm to fit additive models. Different criteria based on the effective degrees of freedom are proposed to choose the penalization parameters. Performance is illustrated with some examples.