

# Bayesian generalized linear models and an appropriate default prior

Andrew Gelman  
Columbia University

Many statistical methods of all sorts have tuning parameters. How can default settings for such parameters be chosen in a general-purpose computing environment such as R? We consider the example of prior distributions for logistic regression.

Logistic regression is an important statistical method in its own right and also is commonly used as a tool for classification and imputation. The standard implementation of logistic regression in R, `glm()`, uses maximum likelihood and breaks down under separation, a problem that occurs often enough in practice to be a serious concern. Bayesian methods can be used to regularize (stabilize) the estimates, but then the user must choose a prior distribution. We illustrate a new idea, the “weakly informative prior”, and implement it in `bayesglm()`, a slight alteration of the existing R function. We also perform a cross-validation to compare the performance of different prior distributions using a corpus of datasets.