ABSTRACT

Generalized linear models (GLMs) (McCullagh and Nelder, 1989) have become popular tools for analyzing biomedical, engineering and agricultural data because of their ability to handle different types of responses in a unified yet parsimonious and interpretable manner. This dissertation presents a novel semiparametric extension of GLMs that removes the need to (correctly) specify an error distribution for the data, yet, somewhat paradoxically, remains fully within the GLM framework and hence retains the same level of parsimony and interpretability.

In particular, we show that both the mean model and the error distribution can be consistently estimated from data. Furthermore, for large enough sample sizes, inferences on the mean model are unaffected by having to also estimate the error distribution from the data. Some interesting small sample simulations and data analysis examples demonstrate that the method already works well with only moderately small datasets.