Generalized Linear Models on Large Data Sets

Joseph Rickert[1,*], Susan Ranney[1]

1. Revolution Analytics
*Contact author: joseph.rickert@revolutionanalytics.com

Keywords: GLM, Tweedie, R, Big Data

Since their introduction by Nelder and Wedderburn over forty years ago, Generalized Linear Models (GLMs) have been a mainstay of statistical inference. Moreover, while they were originally employed with samples of relatively modest size, GLMs are now being employed with very large data sets in machine learning and data science applications. In this talk, we will briefly review the history of using GLMs in *R*, discuss the issues involved in using GLMs on large data sets, and show examples of various models including logistic regression, Poisson and Tweedie models running on large data sets using the Parallel External Memory algorithms implemented in **`RevoScaleR`** package of Revolution R Enterprise.

References

[1] Nelder, J.A. and R.W.M. Wedderburn. 1972 “Generalized linear models.” *Journal of the Royal Statistical Society*, Series A 135:370--84