

A comparison of numerical optimizers for logistic regression

Thomas P. Minka
(2003; revised 10/21/03)

Logistic regression is a workhorse of statistics and is closely related to methods used in Machine Learning, including the Perceptron and the Support Vector Machine. This note compares eight different algorithms for computing the maximum a-posteriori parameter estimate. A full derivation of each algorithm is given. In particular, a new derivation of Iterative Scaling is given which applies more generally than the conventional one. A new derivation is also given for the Modified Iterative Scaling algorithm of Collins et al (2002). Most of the algorithms operate in the primal space, but can also work in dual space. All algorithms are compared in terms of computational complexity by experiments on large data sets. The fastest algorithms turn out to be conjugate gradient ascent and quasi-Newton algorithms, which far outstrip Iterative Scaling and its variants.

[PDF](#)

Matlab code (requires [lightspeed](#)): [logreg.zip](#)

About the experiments: This paper's experiments were done exclusively on synthetic data. Other researchers have investigated whether my findings also hold on real-world data. In every case so far, the real-world studies come to the same conclusions as my synthetic study.

Some real-world implementations of the ideas in this paper (including code):

- [Notes on CG and LM-BFGS Optimization of Logistic Regression](#) by Hal Daume
- [A comparison of algorithms for maximum entropy parameter estimation](#) by Robert Malouf

An earlier version of this paper:

Algorithms for maximum-likelihood logistic regression

Thomas P. Minka
[CMU Statistics Tech Report 758](#) (2001; revised 9/19/03)

Logistic regression is a workhorse of statistics and is increasingly popular in machine learning, due to its similarity with the Support Vector Machine. This note reviews seven different algorithms for finding the maximum-likelihood estimate. Iterative Scaling is shown to apply under weaker conditions than usually assumed. A modified iterative scaling algorithm is also derived, which is equivalent to the algorithm of Collins et al (2000). The best performers in terms of running time are the line search algorithms and Newton-type algorithms, which far outstrip Iterative Scaling.

Equation (11) has been fixed in this version.

[PDF](#) (370K)

[Tom Minka](#)

Last modified: Thu Mar 22 14:50:38 GMT 2007

