An Interior-Point Method for Large-Scale l1-Regularized Logistic Regression

K. Koh, S.-J. Kim, and S. Boyd

Journal of Machine Learning Research, 8:1519-1555, July 2007.
Shorter version appeared as *A method for large-scale l1-regularized logistic regression*, in *22nd National Conference on Artificial Intelligence (AAAI-07)*, 2007.

- l1_logistic_reg.pdf
- l1_logistic_reg_aaai.pdf
- l1_logreg software package

Logistic regression with l1 regularization has been proposed as a promising method for feature selection in classification problems. In this paper we describe an efficient interior-point method for solving large-scale l1-regularized logistic regression problems. Small problems with up to a thousand or so features and examples can be solved in seconds on a PC; medium sized problems, with tens of thousands of features and examples, can be solved in tens of seconds (assuming some sparsity in the data). A variation on the basic method, that uses a preconditioned conjugate gradient method to compute the search step, can solve very large problems, with a million features and examples (*e.g.*, the 20 Newsgroups data set), in a few tens of minutes, on a PC. Using warm-start techniques, a good approximation of the entire regularization path can be computed much more efficiently than by solving a family of problems independently.