Szczecin, 19.09.2024 r.

PhD thesis' abstract

Learning Regularized Linear Models for Small Datasets

Author: MSc Jacek Klimaszewski

Supervisor: Dr. Sc. Marcin Korzeń, prof ZUT

This PhD thesis deals with issue of training regularized linear models on small datasets (i.e. datasets containing much less observations than features). Existing optimization algorithms were analysed and their behaviour on those datasets was examined. Some numerical issues were checked, such as influence of single precision or double precision floating-point type or line-search procedure on training speed. It was proved theoretically and empirically, that in some configurations use of Woodbury's formula or QR factorization lead to shorter time of training ℓ^2-regularized models — the same approach was checked for sparsity-inducing penalizers. Experiments were conducted mainly in Python environment, with some parts written in C++, taking as a reference point existing machine learning libraries, i.e. `scikit-learn` or `liblinear`. Performance of proposed solutions were tested in classification task for some real and sythetic datasets. Moreover, some potential use of training penalized linear models in non-linear problems was highlighted.

keywords: machine learning, linear models, regularization, optimization, Newton's method

19.09.2024

Jacek Klimaszewski