Implicit differentiation of Lasso-type models for hyperparameter optimization

Quentin Bertrand, Parietal Inria Saclay

Abstract

Setting regularization parameters for Lasso-type estimators is notoriously difficult, though crucial in practice. The most popular hyperparameter optimization approach is grid-search using held-out validation data. Grid-search however requires to choose a predefined grid for each parameter, which scales exponentially in the number of parameters. Another approach is to cast hyperparameter optimization as a bi-level optimization problem, that one can solve by gradient descent. The key challenge for these methods is the estimation of the gradient w.r.t. the hyperparameters. Computing this gradient via forward or backward automatic differentiation is possible yet usually suffers from high memory consumption. In addition, implicit differentiation usually assumes smooth loss functions, which is not the case for Lasso-type problems. This work introduces an efficient implicit differentiation algorithm tailored for Lasso-type problems. Our approach scales to high-dimensional data by leveraging the sparsity of the solutions.