Elitenetzwerk Bayern

UNIVERSITÄT BAYREUTH

Lecture Series as part of the elite degree programme Scientific Computing

Date: 27.11.2019 | Time: 4:30 pm | Location: H31, Building FAN B

Coffee / tea from 4:00 pm in the K6, FAN B

www.scientific-computing.uni-bayreuth.de

Regularization Parameter Tracking in Machine Learning

Christopher Schneider

Ernst-Abbe-Hochschule Jena

Abstract: Regularized loss minimization, where a statistical model is obtained from minimizing the sum of a loss function ℓ and weighted regularization terms r_i, is still in widespread use in machine learning:

$$\min_{x \in \mathcal{F}} \ \ell(x) + \sum_{i=1}^{q} \alpha_i r_i(x) \, . \tag{P$_\alpha$}$$

The statistical performance of the resulting models depends on the choice of weights α_i (regularization parameters) that are typically tuned by cross-validation. For finding the best regularization parameters, the regularized minimization problem (P_α) needs to be solved for the whole parameter domain. A practically more feasible approach is covering the parameter domain with approximate solutions of the loss minimization problem for some prescribed approximation accuracy. The problem of computing such a covering is known as the *approximate solution gamut problem.*

Existing algorithms for the solution gamut problem suffer from several problems. For instance, they require a grid on the parameter domain whose spacing is difficult to determine in practice, and they are not generic in the sense that they rely on problem specific plug-in functions. Here, we show that a well-known algorithm from vector optimization, namely *Benson's algorithm*, can be used directly for computing approximate solution gamuts while avoiding the problems of existing algorithms.

Experiments on real world data sets demonstrate the effectiveness of Benson's algorithm for regularization parameter tracking. Therefore, we study

(1) *Linear Regression by the Elastic Net:* In this regression model, a least squares term serves as cost function, while an L^1-regularization term promotes the *sparsity* of the solution and a squared L^2-regularization terms aims for a *grouped* selection of the variables.

(2) *Latent Variable Graphical Models:* Here, a precision matrix for a graphical model has to be decomposed into a *sparse* and a *low-rank* component. Those both structures are obtained by an L^1-norm and a nuclear norm, respectively, which are used as regularization functions. To learn this model, semidefinite programs have to be solved.