Optimization Methods for Machine Learning

Sebastian Sager

Institut für Mathematische Optimierung
Wintersemester

Machine Learning and Optimization

Main questions in this lecture

- How do optimization problems arise in machine learning applications and what makes them challenging?
- What have been the most successful optimization methods for large-scale machine learning and why?
- What recent advances have been made in the design of algorithms and what are open questions in this research area?

Machine Learning and Optimization

Main questions in this lecture

- How do optimization problems arise in machine learning applications and what makes them challenging?
- What have been the most successful optimization methods for large-scale machine learning and why?
- What recent advances have been made in the design of algorithms and what are open questions in this research area?

Additional questions (in the exercises)

- How to use ML software?
- How to implement optimization algorithms for ML?

- What is Machine Learning? Artificial Intelligence?
- How may / will / shall ML and AI change our future?

General ML Optimization Problem

We learn a prediction function $h : \mathcal{X} \mapsto \mathcal{Y}$ given labeled training data $(x_i, y_i)_{i \in [n]}$ with $x_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$:

$$\min_{h \in \mathcal{H}} \quad \underbrace{\frac{1}{n} \sum_{i=1}^{n} \ell(h(x_i), y_i)}_{\text{empirical risk, data fit}} + \underbrace{\Omega(h)}_{\text{regularization}} \tag{0.1}$$

The loss function $\ell : \mathbb{R}^2 \mapsto \mathbb{R}$, the label space $\mathcal{Y}$, the prediction function space $\mathcal{H}$, and the regularization $\Omega : \mathcal{H} \mapsto \mathbb{R}_0^+$ are chosen problem–specificly based on empirical experience.

General ML Optimization Problem

We learn a prediction function $h : \mathcal{X} \mapsto \mathcal{Y}$ given labeled training data $(x_i, y_i)_{i \in [n]}$ with $x_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$:

$$\min_{h \in \mathcal{H}} \quad \underbrace{\frac{1}{n} \sum_{i=1}^{n} \ell(h(x_i), y_i)}_{\text{empirical risk, data fit}} + \underbrace{\Omega(h)}_{\text{regularization}} \tag{0.1}$$

The loss function $\ell : \mathbb{R}^2 \mapsto \mathbb{R}$, the label space $\mathcal{Y}$, the prediction function space $\mathcal{H}$, and the regularization $\Omega : \mathcal{H} \mapsto \mathbb{R}_0^+$ are chosen problem–specificly based on empirical experience.

The chosen combination results in

- a particular approach in ML (SVM, neural networks, ...)
- mathematical properties that can/should be taken into account

Table of Contents

Optimierungsvorlesungen

Vorlesungszyklus Sebastian Sager:

Wann	Titel	SWS	Zielgruppe
WS	Einführung in die Optimierung	4+2	B3
WS	Nichtlineare Optimierung	4+2	B5, M1
SS	G.-g. nichtlineare Optimierung	3+1	B4,B6,M2
WS	Optimization Methods for Machine Learning	4+2	M1
SS	Algorithmische Dynamische Optimierung	3+1	M2

Further lectures on Machine Learning (usually SS):

- Kaibel: Discrete Aspects of Artificial Intelligence
- Richter: Numerik von Differentialgleichungen mit Neuronalen Netzen

Hints

Ideally, you have

- Einführung in die Optimierung
- Nichtlineare Optimierung
- Proficiency in programming (julia, python, ...)

Target group

- Master students mathematics
- PhD students
- Other disciplines welcome, but warning: mathematical point of view!

Comments

- Master thesis possible
- Possibly a seminar OMML next SS
- Online material (videos) available