Convex Optimization for Machine Learning and Computer Vision

Lecture: Dr. Virginia Estellers
Exercises: Emanuel Laude
Winter Semester 2017/18

Computer Vision Group
Institut für Informatik
Technische Universität München

Weekly Exercises 8

Room: 02.09.023
Friday, 26.01.2018, 09:15-11:00
Submission deadline: Monday, 22.01.2018, 10:15, Room 02.09.023

Theory: Stochastic Gradient Descent (8 Points)

Exercise 1 (8 Points)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable and c-strongly convex. Recall that this means that

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle + \frac{c}{2}\|x - y\|_2^2. \tag{1}$$

Prove that f has a unique minimizer x^* and show the following inequality:

$$2c(f(x) - f(x^*)) \leq \|\nabla f(x)\|_2^2.$$

Coding: Variance Reduction (16 Points)

In this exercise you are asked to implement a simple variance reduction approach for SGD, which is also referred to as big-batch stochastic gradient descent. The key idea is to adaptively increase the batch size during iterations.

- Download the solution for full batch proximal gradient descent for logistic regression and alter the code accordingly.
- Increase the batch size during iterations adaptively.
- You may keep the learning rate constant during iterations.
- Argue why a small batch size gives a good descent direction in the beginning but in the limit there is almost no progress.