Convex Optimization for Computer Vision
Lecture: M. Möller
Exercises: J. Geiping
Summer Semester 2017

Universität Siegen
Department ETI
Visual Scene Analysis

Weekly Exercises 6

Room: HF-115

Wednessday, 07.06.2017, 12:15-14:00,

Submission deadline: Monday, 05.06.2017, 12:15, in the lecture

Theory

Exercise 1 (2 points)**.** Let $G : \mathbb{R}^n \to \mathbb{R}^m$ and $F : \mathbb{R}^m \to \mathbb{R}^r$. Show that if one of the operators is a contraction and the other one is non-expansive, then $(F \circ G)$ is a contraction, too.

Exercise 2 (4 points)**.** Let $C \subset \mathbb{R}^n$ be a nonempty, closed convex set. Show that $E_v : \mathbb{R}^n \to \bar{\mathbb{R}}$ defined by

$$E_v(u) = \begin{cases} \|u - v\|^2 & \text{if } u \in C, \\ \infty & \text{otherwise,} \end{cases}$$

has a closed epigraph.

Exercise 3 (4 points)**.** As discussed in the lecture, a classical task in machine learning is *multinomal logistic regression.* Consider some training data (X, t) of n feature vectors of length m, $X \in \mathbb{R}^{n \times m}$, and a desirable *label* $t \in \mathbb{R}^n$ with $t_i = k$ if the i-th example belongs to class $k \in \{1, \dots, c\}$. One seeks to find *weights* $W \in \mathbb{R}^{m,c}$ and *biases* $b \in \mathbb{R}^c$ which minimize the following energy

$$E(W, b) = \quad \frac{1}{n} \sum_{i=1}^{n} l(W, b, X_{i,:}, t_i) + \frac{\lambda}{2} \|W\|_F^2 + \frac{\lambda}{2} \|b\|_2^2. \tag{1}$$

The loss function l is given by

$$\ell(W, b, x, t) = -\log \left(\frac{\exp(xW_{:,t} + b_t)}{\sum_{j=1}^{c} \exp(xW_{:,j} + b_j)} \right) \tag{2}$$

Determine the gradient of E with respect to W and b! (Since this gradient is crucial for the programming exercise, please contact Jonas if you get stuck!)

Programming: Multinomial logistic regression

Exercise 4 (8 Points)**.** Your task is to minimize the energy given in (1) on an example classification problem, namely classifying wines by winery based on its chemical characteristics. As explained at `https://de.mathworks.com/help/nnet/examples/wine-classification.html` the MATLAB *wine_ dataset* contains 178 examples of wines from three different wineries that have been classified with respect to 13 different attributes. Your task is to find weights $W \in \mathbb{W}^{13\times 3}$ and biases $b \in \mathbb{W}^{1\times 3}$ that minimize (1). We will hold back 10% of the overall data to test up to which accuracy you can assign new wines to the respective wineries they came from.

Compare gradient descent methods with backtracking linesearch and with a fixed step size. For the latter you may experimentally determine a step size τ at which the algorithm stabilizes.