TU Bergakademie Freiberg
Institut für Numerische Mathematik und Optimierung
Prof. Dr. O. Rheinbach
Dipl.-Math. Matthias Brändel

Wintersemester 2021/22
16.11.2021

Lineare Algebra, Datenanalyse und maschinelles Lernen 2

4. Exercise Sheet

Exercise 44
Construct data points that are separable by a quadratic function. Learn the decision boundary.

Exercise 45
Contrary to the (batch) gradient descent, the stochastic gradient descent computes the gradient from subset of the available data.

$$\theta^{k+1} = \theta^k + \alpha d^k$$
$$d^k = \sum_{i \in I} \nabla_{\theta_i} f(x_i, y_i, \theta_i)$$

Compare the convergence of the weights and the loss for

1. batch gradient descent
2. mini-batch/stochastic gradient descent for different batch sizes
3. batch gradient descent reusing old d^k (Adam)

   $$d^k = \beta d^{k-1} + (1-\beta)\nabla f$$

4. normalization of α using $\|d^k\|_2$
5. component-wise normalization of α_i

As an example use the minimization problem

$$\min_{x_1,x_2} f(x_1, x_2) = \min_{x_1,x_2} \left(\max(0, \tanh(4x_1 + 4x_2)) + \operatorname{abs}(0.4x_1) + 1\right)$$

with starting values $(2,2)$, $(2,3)$, $(3,2)$.

Exercise 46
Apply the different optimization algorithms from exercise 45 to the MNIST dataset of handwritten digits. You can obtain the data via `tensorflow` using

```
from tensorflow.examples.tutorials.mnist import input_data
mnist = input_data.read_data_sets("MNIST_data/", one_hot=True)
```

Exercise 47
Compute the gradient for a neuronal network with a single neuron.