

Deep Learning

Exercise Sheet 4

Dr. Josif Grabocka, Nicolas Schilling
Information Systems and Machine Learning Lab
University of Hildesheim

May 24th, 2017

Submission until June 1st, 18.00 via Learnweb (id: 3108)

Exercise 7: Optimizing the Rosenbrock Function (10 Points)

- a) In your own words, explain the fundamental difference between normal gradient descent and AdaGrad.
- b) The Rosenbrock function is defined as:

$$z(x, y) = (a - x)^2 + b(y - x^2)^2$$

Compute the partial derivatives of z for arbitrary a and b !

- c) Implement an optimization of the Rosenbrock function for $a = 1$ and $b = 10$ using regular gradient descent and AdaGrad. As starting point, use $x = -1$ and $y = 1$ as well as a learning rate of $\alpha = 1$. Which method converges to the true minimum, which is the point (a, a^2) ?

Exercise 8: Solving the XOR Problem Again (10 Points)

Implement AdaGrad, as well as the normal momentum and the Nesterov's momentum in your implementation of last week's exercise and learn the network again. Which one performs best in your experience?

x_1	x_2	y
1	1	1
-1	1	0
1	-1	0
-1	-1	1