November 28, 2025

Faculty of Computer Science, Heinrich-Heine-University Düsseldorf

Deep Learning

Winter term 25/26 – Exercise Sheet 7

Submission Deadline: Monday, December 8, 2025, 2:00 PM

1. **Momentum** (4P)
   a) Execute two iterations of gradient descent with momentum to find the minimum of the function $f(\theta) = \frac{\theta^3}{3} + 20\theta^2 - 50\theta - 5$. Use a learning rate of $\alpha = 0.01$, momentum $\varphi = 0.5$ and an initial velocity of $v_0 = 0$ and a starting point $\theta_0 = 20$. (2P)
   b) Evaluate the benefit of using a momentum for the task in a) by comparing your findings to the vanilla gradient descent method. (2P)

2. **Adam** (4P)
   Assume the gradient estimate of the fist and second iteration of ADAM are given by $(4, 2, 3, 1)^T$ and $(1, 2, 1, 2)^T$, respectively. Let $\rho_1 = 0,9$ $\rho_2 = 0,99$, and $\alpha = 0.01$. Calculate the first two update steps $\Delta\theta_0$ and $\Delta\theta_1$. Put s_0 and r_0 to 0.

3. **Initialization** (2P)
   Assume you initialize all parameters of a neural network with a small constant c. Is this a good idea? What are better strategies?