Winter term 2024/25
U. von Luxburg
E. Günther/ K. Frohnapfel

Presence sheet 08

Mathematics for Machine Learning

Tutorial of **Week 09 (09.12. - 13.12.2024)**

Exercise 1 (Convergence of Gradient Descent).
Consider the function $f : \mathbb{R} \to \mathbb{R},\ x \mapsto x^2$ and some starting points $x_0 \neq 0$. For which (constant) stepsizes $\eta \in \mathbb{R}$ does gradient descent converge to the minimum? What happens for the other possible step sizes?

Exercise 2 (Optimization in different norms).
Consider the optimization problem

$$\min_{x \in \mathbb{R}} \|x\mathbb{1} - b\|,$$

where $\mathbb{1} \in \mathbb{R}^d$ refers to the vector only consisting of ones and $b \in \mathbb{R}^d$. What is the optimal solution for $x \in \mathbb{R}$ w.r.t. the $\mathcal{L}^1$, the $\mathcal{L}^2$ and the $\mathcal{L}^\infty$-norm?

Exercise 3 (Convergence of Gradient Descent).
Consider the function

$$f : \mathbb{R} \to \mathbb{R},\ x \mapsto \log\left(1 + x^6\right).$$

a) Find the minimum x^* of f analytically.

b) For the starting point $x_0 = 0.2$ and stepsize $\eta = 0.1$, prove that after $k = 500$ steps of gradient descent we still have $|x_k - x^*| > 0.1$.

c) What is the reason for the slow convergence of gradient descent here?

Exercise 4 (Gradient Descent).
Consider the function

$$f : \mathbb{R}^2 \to \mathbb{R},\ (x, y) \mapsto 3x^2 + 4y^2 - 12x - 16y + 10.$$

Apply 4 steps of gradient descent by hand for the starting point $x_0 = (0, 0)$ and the stepsize $\eta = 0.1$.