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

Assignment 07

Mathematics for Machine Learning

Submission due Friday **06.12.24, 23:59** via Ilias

Justify all your claims.

Exercise 1 (**Newton method vs. Gradient Descent**, 1 + 3 + 2 points).
Consider the quadratic optimization problem

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} x^t Q x - b^t x$$

for $Q \in \mathbb{R}^{n \times n}$ symmetric and positive definite and $b \in \mathbb{R}^n$.

a) Show that the optimal solution is $x^* = Q^{-1}b$.

b) To solve the optimization problem iteratively we can use Gradient Descent (GD), which updates the optimal solution in each step by

$$x_{k+1} = x_k - \alpha \nabla f(x_k)$$

for some step size α. Show that for stepsize $\alpha^* = \frac{2}{\lambda_{\min} + \lambda_{\max}}$ the error converges with

$$\|x_{k+1} - x^*\| \leq \left(\frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}} \right) \|x_k - x^*\|$$

where $\lambda_{\max}$ and $\lambda_{\min}$ are the maximal and minimal eigenvalues of matrix Q.
Hint: Use the inequality $\|Av\| \leq \|A\|_2 \cdot \|v\|$ for an arbitrary vector $v \in \mathbb{R}^n$. You don't need to show it.

c) Another possibility to solve the problem is using the Newton method which updates the optimal solution in each step by

$$x_{k+1} = x_k - H^{-1} \nabla f(x_k)$$

where H denotes the Hessian matrix. Show that the Newton method converges to the optimal solution in one step irregardless of the choice of x_0.

Exercise 2 (**Convexity and Continuity**, 1+3+1 points).
Consider a convex function $f : [a, b] \to \mathbb{R}$ for $a < b$, $a, b \in \mathbb{R}$.

a) Prove that for any $x, y, z \in [a, b]$ with $x < y < z$ it holds that

$$\frac{f(y) - f(x)}{y - x} \leq \frac{f(z) - f(x)}{z - x} \leq \frac{f(z) - f(y)}{z - y}.$$

What does this mean geometrically?

b) Show that f is continuous on (a, b).
Hint: Use a) to bound the slope of the line segment between x_0 and an arbitrary point x

c) Is f also continuous in a and b? Give a proof or find a counterexample!

Exercise 3 (**Convexity**, 2+2 points).
Consider a function $f : S \to \mathbb{R}$ on a non-empty convex set $S \subseteq \mathbb{R}^d$. Prove the following statements:

a) f is convex if and only if

$$f\left(\sum_{i=1}^{n} \lambda_i x_i\right) \leq \sum_{i=1}^{n} \lambda_i f(x_i)$$

for all $n \geq 2$, $x_1, \ldots, x_n \in S$ and $0 \leq \lambda_i$, $i \in \{1, \ldots, n\}$ with $\sum_{i=1}^{n} \lambda_i = 1$.

b) For positive $x_1, \ldots, x_n$ the following inequality holds

$$\frac{1}{n}\sum_{i=1}^{n} x_i \geq \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}.$$

Hint: Think of a concave function that might help here. Remember to show that it is concave.

Exercise 4 (**Dual problem**, 3+2 points).

a) Use the method of Lagrange multipliers to solve the following problem.

$$\max x + y$$
$$\text{subject to } x^2 + 2y^2 \leq 5$$

Is the constraint active? Do you have a geometrical explanation of why the constraint should be active or inactive?

b) Consider the linear program

$$\max_{x \in \mathbb{R}^d} c^t x$$
$$\text{subject to } Ax - b \leq 0$$

for $c \in \mathbb{R}^d$, $A \in \mathbb{R}^{m \times d}$ and $b \in \mathbb{R}^m$. Show that the dual of a linear program is again a linear program.
Hint: Derive the dual problem. When is it feasible?