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

Assignment 06

Mathematics for Machine Learning

Submission due Friday **29.11.24, 23:59** via Ilias

Justify all your claims.

Exercise 1 (**Extremal points**, 2+1+2 points).
Consider the function $f\colon \mathbb{R}^2 \to \mathbb{R}, (x, y) \mapsto x^3 + 1/3y^3 - 12x - y$.

a) Compute the set of critical points for f and classify them into local minima, local maxima, or saddle point.

b) Does f have a global minimum or global maximum?

c) Consider the function $g\colon \mathbb{R}^3 \to \mathbb{R}, (x, y, z) \mapsto \alpha x^2 e^y + y^2 e^z + z^2 e^x$ with $\alpha \in \mathbb{R}$. For which values of α is $(0, 0, 0)$ a local minimum, local maximum, or saddle point?

Exercise 2 (**Derivatives**, 2+3 points).
For the following functions, check whether all the directional derivatives exist in $(0, 0)$ and whether the total derivative exists:

a) $f : \mathbb{R}^2 \to \mathbb{R}^2, \quad (x_1, x_2) \mapsto \begin{cases} \frac{x_1 x_2}{\sqrt{x_1^2+x_2^2}} \sin\left(\frac{1}{\sqrt{x_1^2+x_2^2}}\right), & (x_1, x_2) \neq 0 \\ 0, & (x_1, x_2) = 0 \end{cases}.$

b) $g : \mathbb{R}^2 \to \mathbb{R}^2, \quad (x_1, x_2) \mapsto \begin{cases} \frac{x_1^3 x_2}{x_1^4+x_2^2}, & (x_1, x_2) \neq 0 \\ 0, & (x_1, x_2) = 0 \end{cases}.$

Exercise 3 (**Taylor series**, 3 points).
The multivariate Taylor series of a smooth function $f : \mathbb{R}^d \to \mathbb{R}$ around some point $p \in \mathbb{R}^d$ is given by

$$\sum_{n=0}^{\infty} \sum_{\substack{\alpha \in \mathbb{N}_0^d \\ \alpha_1 + \ldots + \alpha_d = n}} \frac{1}{\alpha_1! \cdots \alpha_d!} \cdot \frac{\partial^n f}{(\partial x_1)^{\alpha_1} \cdots (\partial x_d)^{\alpha_d}}(p) \cdot (x_1 - p_1)^{\alpha_1} \cdots (x_d - p_d)^{\alpha_d}.$$

The expression $\frac{\partial^n f}{(\partial x_1)^{\alpha_1} \cdots (\partial x_d)^{\alpha_d}}$ means that for each $i = 1, ..., d$ we take the i-th partial derivative α_i-many times. Note that the order does not matter.
Compute the Taylor series of

$$f : \mathbb{R}^2 \to \mathbb{R},\ (x_1, x_2) \mapsto \frac{1}{1 - x_1 - x_2}$$

around $(0, 0)$ and find the maximal set on which it converges.

Exercise 4 (**Matrix Cookbook**, 2+3+2 points)**.**

a) Show that $\frac{\partial a^t x}{\partial x} = a$ and $\frac{\partial x^t A x}{\partial x} = (A + A^t)x$ for $a \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times n}$ and $x \in \mathbb{R}^n$.

b) In Ridge Regression we want to find optimal weights w^* to fit the linear system $Xw^* = Y$. For that we solve the following optimization problem

$$w^* := \arg\min_{w \in \mathbb{R}^d} \frac{1}{n}\|Xw - Y\|^2 + \lambda\|w\|^2$$

for a design matrix $X \in \mathbb{R}^{n\times d}$, random variable $Y \in \mathbb{R}^n$ and regularization parameter $\lambda > 0$. Use exercise 4 a) to find a closed form solution for w^*.

c) For two functions $h : \mathbb{R}^n \to \mathbb{R}^d$ and $g : \mathbb{R}^d \to \mathbb{R}^k$ the multidimensional chain rule to compute the derivative of $f = g \circ h$ is given by

$$D(g \circ h)(x) = Dg(h(x)) \cdot Dh(x).$$

Use this to compute the derivative of $f : \mathbb{R}^{n\times n} \to \mathbb{R}$ with $f(X) = \log(\det(X))$.