Important information

- Deadline: We. Nov. 7 (Collected Th. 8:00 am)
- Put your printed or clearly handwritten results into the correct box at Samelsonsonplatz. (box number = 60 + group number). Write your name and matrikel number on the sheet.

Exercise 1 (Linear Regression)**.** In this exercise we will see how multi-dimensional linear regression can be used to perform quadratic regression.

1. (2p) Obtain the file 'tutorial4.dat' from the learnweb [1]. It contains two columns x,y of floating point data ($N = 100$). Our goal is to fit a quadratic function $\hat{y}(x) = ax^2 + bx + c$ to the data. Rewrite the MSE in vectorized form:

$$\mathrm{MSE}(\hat{y}) = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}(x_i)|^2 \stackrel{!}{=} \frac{1}{N}\|y - A\beta\|_2^2 \qquad \beta = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \tag{1}$$

2. (3p) Solve the normal equation $A^\intercal A\beta = A^\intercal y$. Report the optimal values and the corresponding MSE. Plot the data and the estimator.

3. (2p) Obtain the file 'tutorial4_2.dat' from the learnweb [2]. It contains three columns x,y,z of floating point data ($N = 1000$). Our goal is to fit a quadratic function

$$\hat{z}(x, y) = v^\intercal B v \qquad B = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \quad v = \begin{pmatrix} x \\ y \end{pmatrix} \tag{2}$$

to the data. Rewrite the MSE in vectorized form:

$$\mathrm{MSE}(\hat{z}) = \frac{1}{N}\sum_{i=1}^{N} |z_i - \hat{z}(x_i, y_i)|^2 \stackrel{!}{=} \frac{1}{N}\|z - A\beta\|_2^2 \qquad \beta = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \tag{3}$$

4. (3p) Solve the normal equation $A^\intercal A\beta = A^\intercal y$. Report the optimal values and the corresponding MSE. (plot not necessary)

Exercise 2 (Gradient descent)**.** In this exercise we want to study the behaviour of gradient descent on the test function $f(x) = x^4 - 1.3x^3 - 1.95x^2 + 4x + 3.65$. This function is non-negative and has precisely one local (and also global) minimum at $x^* = -1$, $f(x^*) = 0$.

[1] https://www.uni-hildesheim.de/learnweb2018/pluginfile.php/71883/mod_resource/content/0/tutorial4.dat
[2] https://www.uni-hildesheim.de/learnweb2018/pluginfile.php/71884/mod_resource/content/0/tutorial4_2.dat

1. (3p) Perform gradient descent with starting point $x_0 = 2$ and step length $\alpha = 0.1$. How many iterations are needed until the function value drops below 10^{-6}?
2. (3p) Try again with the same step length, but from the starting point $x_0 = 1.5$. Why does it take more iterations to achieve the target accuracy, although the starting point is closer to the minimum?
3. (1p) What happens when the starting point $x_0 = -0.5$ with step-length $\alpha = 0.15$ is chosen?
4. (3p) Repeat 1-3, but with Armijo step-length selection, using $\delta = 0.1$. How many iterations are needed in each case? What happens if δ is chosen too large (e.g. $\delta = 0.9$)?