Machine Learning for Image Analysis, SoSe 2024
Exercise Sheet #1: Linear Regression and Gradient Descent
Due date: April 23, 2024, before 11 am

Problem 1 (Linear regression) (10 pt).
You want to perform linear regression on $m \geq 1$ data points $(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})$. The cost function for a general linear hypothesis $h_\theta(x^{(i)}) = \theta_0 + \theta_1 x^{(i)}$ is given by

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2.$$

From the distribution of your data points, you expect the hypothesis to be a constant function, i.e., you set $\theta_1 = 0$.

(a) (4 pts) Formulate the resulting cost function $J(\theta_0)$ explicitly in terms of θ_0 and prove that

$$\frac{dJ}{d\theta_0}(\theta_0) = \theta_0 - \frac{1}{m} \sum_{i=1}^{m} y^{(i)}.$$

(b) (4 pts) For which value of θ_0 is the cost function $J(\theta_0)$ minimal?

(c) (2 pts) You have found a value θ_0 that minimizes the cost function $J(\theta_0)$ and it turns out that the cost is 0. What does this tell you about the data points $(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})$?

Problem 2 (Gradient Descent) (15 pt).
Consider the following cost function in two weight variables

$$J(\theta_1, \theta_2) = \theta_1^4 - 8\theta_1^2 + 16 + 8\theta_2^2.$$

Assume the learning rate α to be 1.

(a) (3 pts) Perform one step of the Gradient Descent algorithm with initial weights $(\theta_1, \theta_2) = (0, 0)$.

(b) (3 pts) Perform one step of the Gradient Descent algorithm with initial weights $(\theta_1, \theta_2) = (2, 0)$.

(c) (6 pts) Check whether $(0, 0)$ and $(2, 0)$ are local extrema or saddle points.
Hint: Take a look at the Hesse Matrix of J in these points. The Hesse Matrix is defined as

$$H = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} := \begin{bmatrix} \frac{\partial^2 J}{\partial \theta_0^2} & \frac{\partial^2 J}{\partial \theta_0 \partial \theta_1} \\ \frac{\partial^2 J}{\partial \theta_1 \partial \theta_0} & \frac{\partial^2 J}{\partial \theta_1^2} \end{bmatrix}$$

*Moreover, a point is a **saddle point** if $\det(H) < 0$. It is a **local minimum** if $\det(H) > 0$ and $H_{11} > 0$. It is a **local maximum** if $\det(H) > 0$ and $H_{11} < 0$.*
(Note: The Hesse Matrix is not part of the exam.)

(d) (2 pts) Find weights (θ_1, θ_2) that minimize the cost function $J(\theta_1, \theta_2)$. Are those weights the unique solution?

(e) (1 pt) Can the above cost function $J(\theta_1, \theta_2)$ stem from a linear regression problem using squared error sum? Explain why or why not.