Machine Learning for Image Analysis, SoSe 2023
Exercise Sheet #1: Linear Regression & Gradient Descent
Due date: April 25, 2023, before 11 am

Problem 1 (Regression vs Classification) (5 pt).
Depending on the described scenario, model the following cases either as a classification problem or a regression problem. Justify your answers.

(a) (1 pt) Based on historical data of the population of African elephants worldwide you want to estimate the number of African elephants worldwide in 2017.

(b) (1 pt) You want to develop an app that takes as input an elephant image and - based on an elephant image database - tells you whether that image depicts an Asian or African elephant, or no elephant at all.

(c) (1 pt) You need to guess the age of an elephant based on features like origin, size, weight and sex of the elephant.

For the following situations, would you design your problem as a supervised or an unsupervised algorithm? Justify your answers.

(d) (1 pt) Given a database of customer data, you are asked to automatically discover market segments and group customers into different market segments.

(e) (1 pt) Given a database of customer data, you are asked to estimate how much a given customer will spend on an average purchase.

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 not part of the lecture.)

(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.