W3WI DS304 Applied Machine Learning Fundamentals

Exercise Sheet # 1 - Numeric Optimization Techniques

Question 1 EX 2023, modified (Gradients)

Compute the gradients of the following functions:

1. $\begin{cases} f_1 : \mathbb{R}^3 \to \mathbb{R} \\ (x, y, z) \mapsto 3x^2 - 5y^2 + 2z^2 \end{cases}$

2. $\begin{cases} f_2 : \mathbb{R}^2 \to \mathbb{R} \\ (x, y) \mapsto \ln\left(\sqrt{xy^3}\right) \end{cases}$

3. $\begin{cases} f_3 : \mathbb{R}^m \to \mathbb{R} \\ \boldsymbol{x} \mapsto \|\boldsymbol{x}\|^2 \end{cases}$

4. $\begin{cases} f_4 : \mathbb{R}^2 \to \mathbb{R} \\ (x, y) \mapsto \frac{xy}{3x+y^2} \end{cases}$

5. $\begin{cases} f_5 : \mathbb{R}^2 \to \mathbb{R} \\ (x, y) \mapsto \ln\left(e^{-x} + e^{2y}\right) \end{cases}$

6. $\begin{cases} f_6 : \mathbb{R}^2 \to \mathbb{R} \\ (x, y) \mapsto \sqrt{x+y}\sin(xy) \end{cases}$

Hint: $\|\boldsymbol{x}\| := \sqrt{\sum_{j=1}^{m} x_j^2}$ denotes the **Euclidean norm** of the vector $\boldsymbol{x} \in \mathbb{R}^m$.

Question 2 (Gradient descent for the Rosenbrock function)

Let the function

$$\begin{cases} f : \mathbb{R}^2 \to \mathbb{R} \\ (x_1, x_2) \mapsto \left(100 \cdot \left(x_2 - x_1^2\right)^2 + \left(x_1 - 1\right)^2\right) \end{cases}$$

be given. This function is known as the **Rosenbrock function** whose graph is shown in the following figure 1:

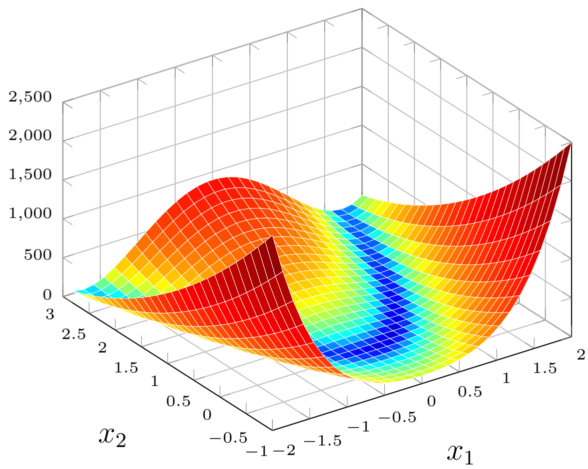

Figure 1: Plot of the two-dimensional Rosenbrock function.

1. Compute the gradient of f and apply five iterations of gradient descent using the learning rate $\alpha = 0.0005$. Start at the point $\boldsymbol{x}_0 = (0.85, 1.10)^\intercal$.

2. Now perform three iterations of gradient descent using the same starting point and the learning rate $\alpha = 0.005$. What phenomenon do you observe?

Question 3 EX 2021, modified (Gradient descent for the Himmelblau function)

Let the following function

$$\begin{cases} f : \mathbb{R}^2 \to \mathbb{R} \\ (x, y) \mapsto \left(x^2 + y - 11\right)^2 + \left(x + y^2 - 7\right)^2 \end{cases}$$

be given (this function is known as the **Himmelblau function**). Perform two iterations of gradient descent using the learning rate $\alpha = 0.02$. Start at the coordinates $x_0 = y_0 = 0$ (the subscript index denotes the number of iterations already performed). Please answer the following questions:

1. What are the values of x_2 and y_2?

2. What is the function value $f(x_2, y_2)$ compared to $f(x_0, y_0)$?

3. Is f a **convex function**? **Hint:** Figure 2 might be helpful.
   (Please justify your answer!)

4. In which point will you eventually end up when initializing the algorithm at the coordinates $\widetilde{x}_0 = -1$ and $\widetilde{y}_0 = 1$?

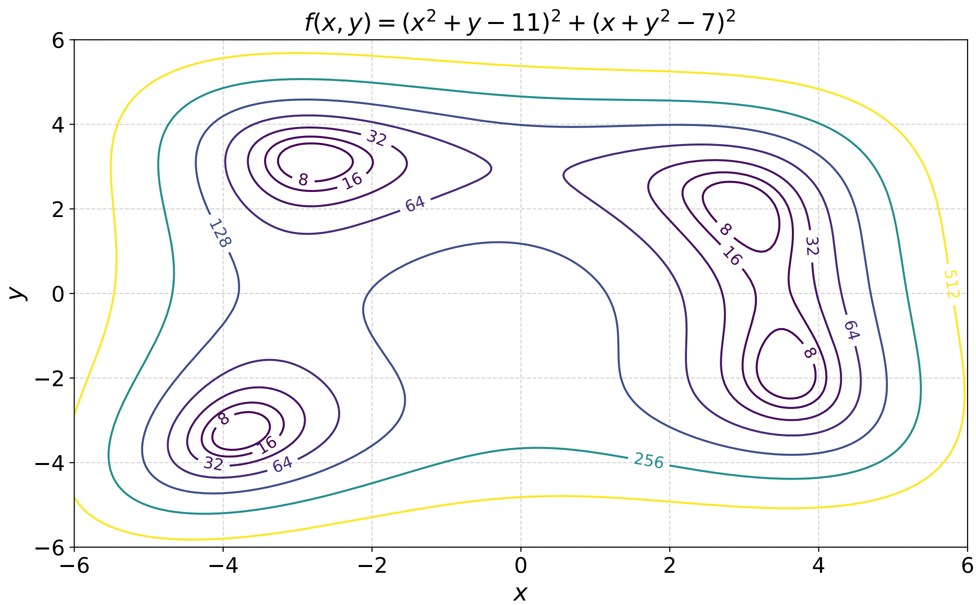

Figure 2: Contour plot of the Himmelblau function.

Question 4 EX 2020 **(Gradient descent learning rate)**

What is a suitable value for the learning rate α? What problems do you face when choosing it **too low or too high**?

Question 5 EX 2021 **(Gradient ascent)**

Suppose you want to find a maximum of a given function f. How would you have to alter the gradient descent update rule to achieve your goal? This algorithm is called **gradient ascent**.

Question 6 EX 2020 **(Gradient descent update rule)**

Tick the correct parameter update rule used in gradient descent ($\mathcal{J}$ is the cost function).

- ☐ $\boldsymbol{\theta} \longleftarrow \boldsymbol{\theta} + \alpha \nabla \mathcal{J}(\boldsymbol{\theta})$
- ☐ $\boldsymbol{\theta} \longleftarrow \boldsymbol{\theta} - \alpha \nabla \mathcal{J}(\boldsymbol{\theta})$
- ☐ $\boldsymbol{\theta} \longleftarrow \alpha \nabla \mathcal{J}(\boldsymbol{\theta})$
- ☐ All options are incorrect.

Question 7 (Newton's method)

Compute the Hessian matrix of the Rosenbrock function (see question 2) and perform two iterations of Newton's method starting from the coordinates $\boldsymbol{x}_0 = (0.85, 1.10)^\intercal$!