Artificial Intelligence and Machine Learning

Exercises – Numeric Optimization Techniques

Question 1 (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. $\begin{cases} f_2 : \mathbb{R}^2 \to \mathbb{R} \\ (x, y) \mapsto \ln(\sqrt{xy^3}) \end{cases}$

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

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

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

Question 2 (Gradient descent for the Rosenbrock function)

Let the function

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

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

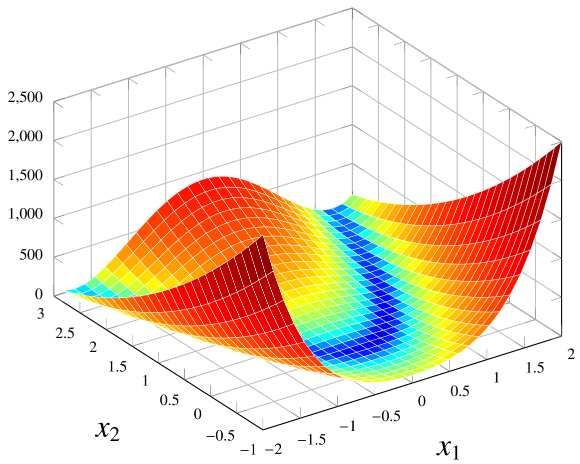

Figure 1: Plot of the two-dimensional Rosenbrock function.

1. Compute the gradient of f and perform five iterations of gradient descent using the learning rate $\alpha := 0.0005$. Start at the point $\boldsymbol{x}^0 = \begin{pmatrix} 0.85 & 1.10 \end{pmatrix}^\intercal$.

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 (Gradient descent for the Himmelblau function) ⊛

Let the following function

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

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 $\boldsymbol{x}^0 := \boldsymbol{0}$ *(zero vector)*. Please answer the following questions:

1. What is the value of $\boldsymbol{x}^2$ *(position after two iterations)*?

2. What is the function value $f(\boldsymbol{x}^2)$ compared to $f(\boldsymbol{x}^0)$?

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

4. In which points will you eventually end up when initializing the algorithm at the coordinates $\boldsymbol{x}^0 := \boldsymbol{0}$ and $\widetilde{\boldsymbol{x}}^0 := \begin{pmatrix} -1, 1 \end{pmatrix}^\intercal$, respectively? Write a Python program and perform a sufficient amount of iterations!

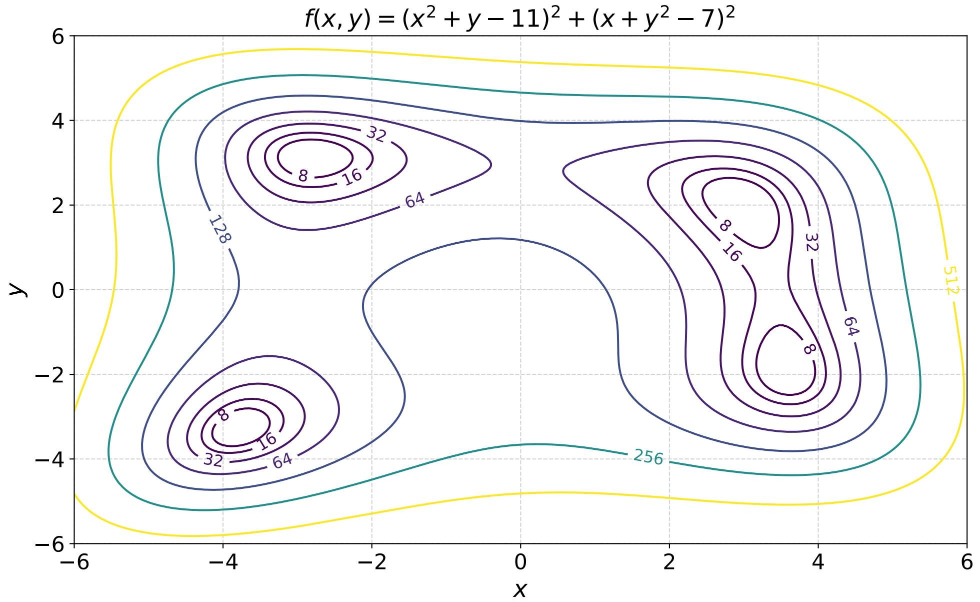

Figure 2: Contour plot of the Himmelblau function.

Question 4 (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 (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?

Remark: The resulting algorithm is called **gradient ascent**.

Question 6 (Gradient descent update rule) ⊛

Tick the correct parameter update rule used in gradient descent for the function f.

- ☐ $\boldsymbol{\theta}^{t+1} \longleftarrow \boldsymbol{\theta}^t + \alpha \nabla f(\boldsymbol{\theta}^t)$
- ☐ $\boldsymbol{\theta}^{t+1} \longleftarrow \boldsymbol{\theta}^t - \alpha \nabla f(\boldsymbol{\theta}^t)$
- ☐ $\boldsymbol{\theta}^{t+1} \longleftarrow \alpha \nabla f(\boldsymbol{\theta}^t)$
- ☐ 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 $\boldsymbol{x}^0 := \begin{pmatrix} 0.85 & 1.10 \end{pmatrix}^\top$.

Question 8 (Constrained optimization)

Let the function

$$f(x_1, x_2) = x_1^2 + \frac{1}{2} x_2^2 + 3$$

be given. Work through the following tasks:

1. Write f in the form
$$f(\boldsymbol{x}) = \frac{1}{2} \boldsymbol{x}^\top \boldsymbol{Q} \boldsymbol{x} + \boldsymbol{c}^\top \boldsymbol{x} + \alpha,$$
where $\boldsymbol{Q} \in \mathbb{R}^{2\times 2}$ is a symmetric matrix, $\boldsymbol{c} \in \mathbb{R}^2$ and $\alpha \in \mathbb{R}$.
Is f a convex function? *(Please justify your answer!)*

2. Compute the (unconstrained) minimum of f analytically.

3. Let the constraint
$$h(x_1, x_2) := x_1 + x_2 - 1 = 0$$
be given. What is the (constrained) solution given this additional constraint?

Question 9 (Dual optimization problem)

Consider the following primal optimization problem

$$\begin{aligned} &\textit{minimize} && \boldsymbol{x}^\top \boldsymbol{x} \\ &\textit{subject to} && \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \end{aligned}$$

where $\boldsymbol{A} \in \mathbb{R}^{K \times M}$. Thus, this minimization problem contains K (linear) equality constraints and no inequality constraints. Please work through the following tasks:

1. Derive the dual optimization problem!

2. Consider the above optimization problem for the case $K = M = 2$. Let the equality constraints be given by

$$\begin{aligned} h_1(x_1, x_2) &:= 2x_1 - 1 = 0 \\ h_2(x_1, x_2) &:= x_1 + x_2 = 0. \end{aligned}$$

   Solve this optimization problem

   (a) via the primal formulation, and

   (b) via the dual formulation.

   Convince yourself that the solutions are indeed identical.