Machine Learning I — Exercise 1

Sacha Sokoloski, Sebastian Damrich – based on material by Dmitry Kobak 16 Oct 2025.
Deadline: 20 Oct 2025 6pm

Task 1 Consider a one-dimensional loss function $\mathcal{L}(\beta) = 3\beta^2$. You are minimizing it with gradient descent starting from some initial β_{init}. For which values of the learning rate η will the loss decrease / increase / remain the same in the first step? Does your answer depend on β_{init}?

Task 2 Consider the same loss function as above. Let k be the gradient update iteration number and β_k the value of β in the k-th iteration. Instead of the function

$$\mathcal{L} : \mathbb{R} \to \mathbb{R}, \beta \mapsto \mathcal{L}(\beta) = 3\beta^2$$

mapping β to its loss value, we can consider the function that maps an update iteration k to a loss value

$$\mathcal{L} : \mathbb{N} \to \mathbb{R} \to \mathbb{R}, k \mapsto \beta_k \mapsto \mathcal{L}(k) = \mathcal{L}(\beta_k).$$

What is shape of the function $k \mapsto \mathcal{L}(k)$ when the learning rate is small enough that the loss decreases? What is the shape if it is too large and the loss increases?
Hint: Possible shapes are, e.g., constant, $1/k$, $1/k^2$, exponential, linear, quadratic etc. Try to write β_k as a direct function of β_{init}, η and k. What kind of function is it? What does this imply for the shape of the loss?

Task 3 Consider a simple linear regression problem with one predictor and the loss function

$$\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2.$$

Derive explicit analytical solutions for $\hat{\beta}_0$ and $\hat{\beta}_1$ minimizing this loss.
Hint: Try to solve the linear system derived from the two partial derivatives.