Exercices 7

TABLE DES MATIÈRES

1 Optimization of the learning rate 1

1 OPTIMIZATION OF THE LEARNING RATE

In this exercise we consider gradient descent for a least-squares problem.

— $\mathcal{X} = \mathbb{R}^d$

— $\mathcal{Y} = \mathbb{R}$

— Design matrix : X

— Outputs : $y \in \mathbb{R}^n$.

We want to minimize the function f representing the empirical risk :

$$f(\theta) = \frac{1}{2n}||X\theta - y||^2 \tag{1}$$

We recall that the gradient and the Hessian write :

$$\begin{aligned} \nabla_\theta f &= \frac{1}{n}X^T(X\theta - y) \\ &= H\theta - \frac{1}{n}X^T y \end{aligned} \tag{2}$$

$$H = \frac{1}{n}X^T X \tag{3}$$

We note the gradient update $\theta_{t+1} = \theta_t - \gamma\nabla_{\theta_t} f$

Considering an fixed iteration step θ_t, we note

$$\alpha(\gamma) = \theta_t - \gamma\nabla_{\theta_t} f \tag{4}$$

Given θ_t, we consider the problem of finding the optimal learning rate, which means the rate γ^* that minimizes the function

$$\begin{aligned} g(\gamma) &= f(\theta_t - \gamma\nabla_{\theta_t} f) \\ &= f(\alpha(\gamma)) \end{aligned} \tag{5}$$

Given θ_t, what is the optimal choice of γ, noted γ^* ?