

DAY 3. INTRO TO MACHINE LEARNING. EMINES.

GRADIENT DESCENT ALGORITHMS

EXERCISE 1 (GRADIENT DESCENT UNDER L -SMOOTH ASSUMPTION) Let f be convex and L -smooth on $\mathbb{R}^d$. We assume that the minimum of f is attained at x^* . Then, we will show that the gradient descent algorithm with $\eta = \frac{1}{L}$ satisfies

$$f(x_t) - f(x^*) \leq \frac{2L \|x_1 - x^*\|}{t-1}$$

We first assume that we are able to prove that $\|x_s - x^*\|$ is decreasing.

1. Show that $f(x_{s+1}) - f(x_s) \leq -\frac{1}{2L} \|\nabla f(x_s)\|^2$.
2. Define $\delta_s = f(x_s) - f(x^*)$. Show that $\delta_s \leq \|x_s - x^*\| \cdot \|\nabla f(x_s)\|$.
3. Deduce that $\delta_{s+1} - \delta_s \leq -\frac{1}{2L\|x_1 - x^*\|^2} \delta_s^2$.
4. Deduce that $-\frac{1}{\delta_{s+1}} + \frac{1}{\delta_s} \leq -\frac{1}{2L\|x_1 - x^*\|^2}$ and conclude.

We now show that $\|x_s - x^*\|$ is decreasing.

5. Show that for any $x, y \in \mathbb{R}^d$, $f(x) - f(y) \leq \langle \nabla f(x), x - y \rangle - \frac{1}{2L} \|\nabla f(x) - \nabla f(y)\|^2$.
6. Using this inequality with $x = x_s$ and $y = x^*$, deduce that $\|x_s - x^*\|$ is decreasing

EXERCISE 2 (PRELIMINARIES ON CONVEX ANALYSIS) Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be a convex function and $\mathcal{C} \subset \mathbb{R}^d$ be a closed convex set. Denote by $\pi_{\mathcal{C}}x = \operatorname{argmin}_{u \in \mathcal{C}} \|x - u\|^2$ the projection of x onto the convex set $\mathcal{C}$.

1. Prove that for all $u \in \mathcal{C}$ and $0 < t < 1$, $\|z - (tu + (1-t)\pi_{\mathcal{C}}z)\|^2 \geq \|z - \pi_{\mathcal{C}}z\|^2$?
2. Deduce from the previous question that

$$\langle u - \pi_{\mathcal{C}}z, z - \pi_{\mathcal{C}}z \rangle \leq 0 \quad \text{and} \quad \|\pi_{\mathcal{C}}z - z\|^2 + \|u - \pi_{\mathcal{C}}z\|^2 \leq \|u - z\|^2.$$

3. Assume that f is differentiable and convex. For any $x, h \in \mathbb{R}^d$ and $t \in [0, 1]$, define

$$F(t) = f(x + th).$$

Prove that $F(1) - F(0) \geq F'(0)$ and conclude that for all $x, y \in \mathbb{R}^d$,

$$f(y) - f(x) \geq \langle \nabla f(x), y - x \rangle.$$

EXERCISE 3 (CONVERGENCE RATES FOR STRONGLY CONVEX FUNCTIONS) When the function f is strongly convex, then the PGD converges much faster. Assume that f is α -strongly convex :

$$f(x) - f(y) \leq \langle \nabla f(x), x - y \rangle - \frac{\alpha}{2} \|x - y\|^2 \tag{1}$$

and that ∇f is β -Lipschitz. The aim of this section is to prove that, for $\eta = 1/\beta$,

$$\|x_{K+1} - x^*\|^2 \leq \|x_1 - x^*\|^2 e^{-\rho K},$$

with $\rho = \alpha/\beta$. Define

$$g : x \mapsto \beta \left(x - \pi_{\mathcal{C}} \left(x - \frac{1}{\beta} \nabla f(x) \right) \right).$$

The key of the proof is to obtain that, for all $(x, y) \in \mathcal{C}^2$,

$$f(x^+) - f(y) \leq \langle g(x), x - y \rangle - \frac{1}{2\beta} \|g(x)\|^2 - \frac{\alpha}{2} \|x - y\|^2, \tag{2}$$

where $x^+ = \pi_{\mathcal{C}}(x - \beta^{-1} \nabla f(x))$.

1. Assume first that (2) holds. Prove the following (in)equalities :

$$\begin{aligned}\|x_{k+1} - x^*\|^2 &= \|x_k - x^*\|^2 - \frac{2}{\beta} \langle g(x_k), x_k - x^* \rangle + \frac{1}{\beta^2} \|g(x_k)\|^2, \\ &\leq (1 - \rho) \|x_k - x^*\|^2 \leq e^{-\rho k} \|x_1 - x^*\|^2.\end{aligned}$$

2. It remains to prove (2). First, show that

$$f(y) - f(x) = \int_0^1 \langle \nabla f(x + t(y - x)), y - x \rangle dt \leq \langle \nabla f(x), y - x \rangle + \frac{\beta}{2} \|y - x\|^2. \quad (3)$$

3. Remind that $x^+ = \pi_{\mathcal{C}}(x - \frac{1}{\beta} \nabla f(x))$. Using (1) and (3), check that

$$f(x^+) - f(y) \leq \langle \nabla f(x), x^+ - x \rangle + \frac{\beta}{2} \|x^+ - x\|^2 + \langle \nabla f(x), x - y \rangle - \frac{\alpha}{2} \|y - x\|^2.$$

4. With Exercice 2 Question 2, prove that $\langle \nabla f(x), x^+ - y \rangle \leq \langle g(x), x^+ - y \rangle$ for all $y \in \mathcal{C}$.

5. Conclude that

$$\begin{aligned}f(x^+) - f(y) &\leq \langle g(x), x^+ - y \rangle + \frac{1}{2\beta} \|g(x)\|^2 - \frac{\alpha}{2} \|y - x\|^2, \\ &= \langle g(x), x - y \rangle - \frac{1}{2\beta} \|g(x)\|^2 - \frac{\alpha}{2} \|y - x\|^2.\end{aligned}$$