Subgradient and Subdifferential

Seminar

Optimization for ML. Faculty of Computer Science. HSE University

Main notions recap

Main notions

For a domain set $E \in \mathbb{R}^n$ and a function $f : E \to \mathbb{R}$:

- A vector $g \in \mathbb{R}^n$ is called **subgradient** of the function f at $x \in E$ if $\forall y \in E$

$$f(y) \geq f(x) + g^T(y - x)$$

- A set $\partial f(x)$ is called **subdifferential** of the function f at $x \in E$ if:

$$\partial f(x) = \{g \in \mathbb{R}^n \mid f(y) \geq f(x) + g^T(y - x)\} \forall y \in E$$

- $f(\cdot)$ is called **subdifferentiable** at point $x \in E$ if $\partial f(x) \neq \emptyset$

Connection between subdifferentiation and convexity

Connection between subdifferentiation and convexity

If $f : E \to \mathbb{R}$ is subdifferentiable on the **convex** subset $S \in E$ then f is convex on S.

- The inverse is generally incorrect
- There is no sense to derive the subgradient of nonconvex function.

Connection between subdifferentiation and differentiation

Connection between subdifferentiation and differentiation

1) If $f : E \to \mathbb{R}$ is convex and differentiable at $x \in \text{int } E$ then $\partial f(x) = \{\Delta f(x)\}$
2) If $f : E \to \mathbb{R}$ is convex and for $x \in \text{int } E$ $\partial f(x) = \{s\}$ then f is differentiable at x and $\Delta f(x) = s$

- Derive the subdifferencial of a differentiable function is overkill.

Subgradient descent

The subgradient method is a straightforward algorithm for minimizing a convex function that is not differentiable:

$$x_{k+1} = x_k - \alpha_k g_k$$

Here $g_k \in \partial f(x_k)$ and $\alpha_k > 0$. There are several well-known α_k selection strategies for this method:

- $\alpha_k = \alpha$ – fixed step size (for G-Lipshitz functions it can give a constant residual $\frac{G^2\alpha}{2}$ between f^* and f_k)
- $\alpha_k = \dfrac{\gamma}{\|g_k\|_2}$ – constant step length ($\|x_{k+1} - x_k\| = \gamma$)
- $\alpha_k : \sum_{k=1}^{\infty} \alpha_k^2 < \infty$ and $\sum_{k=1}^{\infty} \alpha_k = \infty$ – square summable but not summable
- $\alpha_k : \lim_{k\to\infty} \alpha_k = 0$ and $\sum_{k=1}^{\infty} \alpha_k = \infty$ – nonsummable diminishing
- $\alpha_k = \dfrac{f(x_k) - f^*}{\|g_k\|_2^2}$ – Polyak step size

Problem 1

Question

Find the subgradient of the function

$$f(x) = -\sqrt{x}$$

Subdifferentiation rules

1) $f : E \to \mathbb{R}$, $x \in E$, $c > 0$

$$\Rightarrow \partial(cf)(x) = c\partial f(x)$$

2) $f : F \to \mathbb{R}$, $g : G \to \mathbb{R}$, $x \in F \bigcap G$

$$\Rightarrow \partial(f + g)(x) \supseteq \partial f(x) + \partial g(x)$$

3) $T : V \to W = Ax + b$, $g : W \to \mathbb{R}$, $x_0 \in V$

$$\Rightarrow \partial(g \circ T)(x_0) \supseteq A^* \partial(g)(T(x_0))$$

4) $f(x) = \max(f_1(x), \ldots, f_m(x))$, $I(x) = \{i \in 1 \ldots m | f_i(x) = f(x)\}$

$$\Rightarrow \partial f(x) \supseteq \text{Conv}(\bigcup_{i \in I(x)} \partial f_i(x))$$

When is equality reached?

If abovementioned functions are convex and x is inner point then all inequalities turn into equalities.

Problem 2

Question

Find the subgradient of the function $f(x) + g(x)$ if

$$f(x) = -\sqrt{x} \text{ when } x \geq 0$$

$$g(x) = -\sqrt{-x} \text{ when } x \leq 0$$

Problem 3

Question

1) Find the subdifferential of the function $f(x) = ||Ax - b||_1$;
2) For task $f(x) = \frac{1}{2}||Ax - b||_2^2 + \lambda||x||_1 \rightarrow \min_x$ say which lambdas lead to $x_{opt} = 0$

Problem 4

Exercise

We want to find a vector of weights $w \in \mathbb{R}^d$, defining the linear classifier $sign(w^T x)$. Use SVM with l_2-regularization:

$$L(w) = \frac{\lambda}{2}\|w\|^2 + \frac{1}{N}\sum_{i=1}^{N} \max\left\{0,\ 1 - y_i w^T x_i\right\}$$

Here $\ell_i(w) = \max\left\{0,\ 1 - y_i w^T x_i\right\}$ is a hinge loss. Show that the subgradient at any point is given by:

$$g(w) = \lambda w + \frac{1}{N}\sum_{i=1}^{N} \begin{cases} -y_i x_i, & \text{if } y_i w^T x_i < 1, \\ 0, & \text{otherwise} \end{cases}$$

and check out the sample code here.

Problem 5

Question

Find the subdifferential $\partial f(x)$ of the function $f(x) = exp(|x-1| + |x+1|)$ for all $x \in \mathbb{R}$