Subgradient Method

Chirag Mehta*

Indian Institute of Technology, Hyderabad
ai20btech11006@iith.ac.in

April 23, 2022

Abstract

Non-differentiable functions are an important class of functions which often appear in optimization problems, a gradient descent method would fail to optimize such an objective because the gradient might not exist. Methods such as interior point method work great, but it has its own limitations of being computationally inefficient. We will explore the subgradient method which is an iterative first-order method similar to gradient descent. We will also be exploring the heavy-ball method to make the subgradient method faster.

I. Introduction

Subgradient method is a simple algorithm used to minimize non-differentiable convex functions. This method is similar to vanilla gradient method which is used to optimize differentiable functions.

i. Algorithm

Lets say we have a nondifferentiable function $f : \mathbb{R}^n \to \mathbb{R}$. The update rule of subgradient method says

$$\underline{x}^{(k+1)} = \underline{x}^{(k)} - \alpha_k \underline{g}^{(k)} \tag{1}$$

where $\underline{x}^{(k)}$ is the k^{th} iterate, α_k is the step step size at k^{th} iteration and $\underline{g}^{(k)}$ is any subgradient of f at $\underline{x}^{(k)}$. The subgradient $\underline{g}^{(k)}$ is any vector which satisfies

$$f(\underline{y}) \geq f(\underline{x}) + \underline{g}^T(\underline{y} - \underline{x}) \tag{2}$$

At a given point there can be more than one subgradients, we call the set of subgradients as subdifferential.

Theorem I.1. *If the function f is differentiable at $\underline{x}^{(k)}$ then $\underline{g}^{(k)}$ is equal to the gradient of f at $\underline{x}^{(k)}$*

Proof. Substitute $\underline{y} = \underline{x} + \lambda \underline{z},\ \lambda > 0$ in (**??**)

$$\frac{f(\underline{x} + \lambda \underline{z}) - f(\underline{x})}{\lambda} \geq \underline{g}^T \underline{z} \tag{3}$$

We can use the limit $\lambda \to 0$

$$\underline{\nabla} f(\underline{x})^T \underline{z} \geq \underline{g}^T \underline{z} \tag{4}$$

$$\underline{z}^T \left(\underline{\nabla} f(\underline{x}) - g \right) \geq 0 \quad \forall \underline{z} \tag{5}$$

$$\therefore \underline{g} = \underline{\nabla} f(\underline{x}) \tag{6}$$

□

II. SVM using subgradient method

A support vector machine is used for two class classification. The objective is to maximize the slab thickness while still satisfying few constraints. Hard Margin SVM problem can be formuated as follows

$$\min_{\underline{w}, b} \ \underline{w}^T \underline{w} \tag{7}$$

$$\text{s.t: } y_i(\underline{w}^T \underline{x_i} + b) \geq 1 \tag{8}$$

$$\tag{9}$$

We can transform this problem into

$$\min_{\underline{w}, b} \ \underline{w}^T \underline{w} + \lambda \sum_i \max(0, 1 - y_i) \tag{10}$$

*This is a preliminary report, the main report will be updated at github

III. Convergence proof

Lets assume x^* is the minimizer of our objective function f. Assume that the norm of subgradients is bounded.

Using Lipschitz condition

$$|f(u) - f(v)| \leq G\|u - v\|_2 \tag{11}$$

for all u, v. Some versions of subgradient method work even when the gradient is not bounded.

IV. Preliminary References

- A youtube video on subgradient method
- Stanford notes
- Mathematics behind subgradients
- Subgradient method in SVM
- Subgradient method in SVM
- Heavy-ball method