Constrained Optimization Problem

13 April 2025 12:17

Recap with some changes in notations:

In gradient descent, the core formula was:

$X_{(i+1)} = X_i - h * d/dx f(x)$

But the above formula is valid only for a Loss function with one variable x. For a loss function in two variables, this formula will become like this:

$X_{(i+1)} = X_i - h * [d/dx f(x, y) + d/dy f(x, y)]$

For n number of variables:

$$x_{t+1} = x_t - \eta \left[\frac{\partial}{\partial x_1} f(x_1, x_2, \ldots x_n) + \frac{\partial}{\partial x_2} f(x_1, x_2, \ldots x_n) + \ldots + \frac{\partial}{\partial x_n} f(x_1, x_2, \ldots x_n)\right]$$

$$\therefore x_{t+1} = x_t - \eta \sum_{i=1}^{n} \frac{\partial}{\partial x_i} f$$

Now by replacing x_t by θ_t or θ^t & x_{t+1} by θ_{t+1} or θ^{t+1}:

$$\theta^{t+1} = \theta^t - \eta \sum \frac{\partial}{\partial \theta_i^t} f$$

Constrained Optimization Problem:

As we know, our loss function is- $L = \sum \frac{\vec{\omega}^T \vec{x} + \omega_0}{\|\vec{\omega}\|} \cdot y_i$

While computing gradient, we will need to differentiate this function and the division will add complexity in the differentiation.
Constrained optimization problem is a strategy in which we simplify our problem by adding some constraints (some limitation) so that it becomes easy (less complex) to optimize it.

In this case, $\|\vec{\omega}\|$ in the denominator is increasing the complexity so, let's take a constraint - $\|\vec{\omega}\| = 1$

Now with this constraint, we want to find the values of $\vec{\omega}$ & ω_0 for which the value of Loss function becomes minimum. This can be also written as the following

along with our constraint:

$$\underset{(\vec{\omega},\, \omega_0)}{\operatorname{argmin}} \; -\sum \frac{\vec{\omega}^T \cdot \vec{x}_i + \omega_0}{\|\vec{\omega}\|} \cdot y_i \qquad \underline{\underline{s.t.}} \quad \|\vec{\omega}\| = 1$$

↳ such that / Subject to

Now, let's calculate the gradient ∇L

$$\frac{\partial}{\partial \omega_1} L = -\frac{\partial}{\partial \omega_1} \left(\omega_1 x_1 y_1 + \omega_2 x_2 y_2 + \dots + \omega_n x_n y_n + \omega_0 y_0 \right)$$

$$= -x_1 y_1$$

$$\frac{\partial}{\partial \omega_2} L = -x_2 y_2 \quad \dots \quad \frac{\partial}{\partial \omega_n} L = -x_n y_n$$

$$\nabla L = \frac{\partial}{\partial \omega_1} L + \frac{\partial}{\partial \omega_2} L + \dots + \frac{\partial}{\partial \omega_n} L = -\left(x_1 y_1 + x_2 y_2 + \dots + x_n y_n \right)$$

$$\nabla L = -\sum x_i y_i$$

Now, putting this into gradient descent formula to predict next value of ω

$$\omega_{i+1} = \omega_i - \eta \nabla L$$

$$\boxed{\omega_{i+1} = \omega_i - \eta \left(-\sum x_i y_i \right)}$$

The above formula is valid for all components of $\vec{\omega}$ other than ω_0

For ω_0,

$$\boxed{w_0^{t+1} = w_0^t - \eta\left(-\Sigma y_i\right)}$$

A deeper dive into constrained optimization:

Example - 1:
Suppose there is a function f(x) = x^2 - 3x - 3 that we want to minimize and, we also want to put a constraint that our minima must also satisfy the function g(x) = x^2 - 2x - 3 then our constrained optimized problem can be represented as:

$$\boxed{\underset{x}{\operatorname{argmin}}\ f(x) \quad \text{s.t.} \quad g(x)}$$

The goal here is to minimize the f(x) in a way that the constraint g(x) is also taken care of. But there is a problem with the above equation - how to differentiate a formula with "s.t." part in it? Answer to this question is known as - **Lagrange's Multiplier**.

Instead of differentiating (or working with) the above equation, we will first transform it in the following form:

$$\underset{x,\lambda}{\operatorname{argmin}}\ \underbrace{f(x)}_{\text{Minimizin}} + \underbrace{\lambda\, g(x)}_{\text{Taking care of the constraint}} \qquad \text{I}$$

Now this equation doesn't have the "s.t." part in it hence this form is called "Unconstrained Problem".

This equation I is considering only one constraint. What if we have multiple constraints?

$$\underset{x,\lambda_1,\lambda_2\ldots\lambda_n}{\operatorname{argmin}}\ f(x) + \lambda_1 g_1(x) + \lambda_2 g_2(x) + \ldots\ldots + \lambda_n g_n(x)$$

In the above formula, λ_1, l_2, l_3, … , l_n are called **Lagrange's Multipliers**.

Now let's try to solve the **Example - 1** using this approach:

$$f(x) = x^2 - 3x - 3 \ ,\ g(x) = x^2 - 2x - 3$$

$$\underset{x,\lambda}{\operatorname{argmin}} \; f(x) + \lambda g(x) = \underset{x,\lambda}{\operatorname{argmin}} \; x^2 - 3x - 3 + \lambda x^2 - 2\lambda x - 3\lambda = h(x,\lambda)$$

To minimize this, we need to calculate its gradient:

$$\nabla h = \begin{bmatrix} \frac{\partial}{\partial x} h(x,\lambda) \\ \\ \frac{\partial}{\partial \lambda} h(x,\lambda) \end{bmatrix} = \begin{bmatrix} 2x - 3 + 2\lambda x - 2\lambda \\ \\ x^2 - 2x - 3 \end{bmatrix}$$

At the minima, the differentiation (gradient) is 0.

$\therefore \nabla h = 0$

$\therefore 2x - 3 + 2\lambda x - 2\lambda = 0 \quad \text{(A)}$ & $x^2 - 2x - 3 = 0$

$x^2 - 3x + x - 3 = 0$

$x(x-3) + 1(x-3) = 0$

$\boxed{x = -1}$ OR $\boxed{x = 3}$

By putting $x = -1$ & $x = 3$ into (A):

$\boxed{x = -1:}$

$2(-1) - 3 + 2\lambda(-1) - 2\lambda = 0$

$\therefore -2 - 3 - 2\lambda - 2\lambda = 0$

$\therefore -4\lambda = 5 \Rightarrow \boxed{\lambda = \frac{-5}{4}}$

For $\boxed{x = 3:}$ $2(3) - 3 + 2\lambda(3) - 2\lambda = 0$

$\therefore 3 + 6\lambda - 2\lambda = 0$

$\boxed{\lambda = \frac{-3}{4}}$

$\therefore$ The two minima/maxima are $\left(-1, \frac{-5}{4}\right)$ & $\left(3, \frac{-3}{4}\right)$

To identify the minimum of these two:

$f(x) + \lambda g(x)$ for $x=-1$ & $\lambda = \left(\frac{-5}{4}\right)$:

$$x^2 - 3x - 3 + \lambda x^2 - 2\lambda x - 3\lambda = 1 + \cancel{3} - \cancel{3} - \frac{5}{4} - \frac{2\times 5}{4} + \frac{3\times 5}{4}$$

$$= \frac{4 - 5 - 10 + 15}{4} = \frac{4}{4} = 1 \quad \therefore \boxed{f(x) + \lambda g(x) = 1 \text{ for } x=-1 \ \& \ \lambda = -\frac{5}{4}}$$

similarly for $x = 3$ & $\lambda = -\frac{3}{4}$:

$$\cancel{9} - \cancel{9} - 3 + \frac{-3}{4}\times 9 - 2\times\frac{(-3)}{4}\times 3 - 3\left(\frac{-3}{4}\right) = \frac{-12 - 27 + 18 + 9}{4}$$

$$= \frac{-12}{4} = -3 \quad \boxed{\therefore f(x) + \lambda g(x) = -3 \text{ for } x = 3 \ \& \ \lambda = \frac{-3}{4}}$$

$\therefore$ Our minim is at $\left(3, \frac{-3}{4}\right)$ ___ solution

Now let's apply this approach to minimize the loss function of our G.D.

$$f(x) = L = -\sum \frac{\vec{\omega}^T \cdot \vec{x} + \omega_0}{\|\vec{\omega}\|} \cdot y_i \quad \text{s.t.} \quad \underbrace{\|\vec{\omega}\| = 1}$$

At this place it must be of the form $g(x) = 0$

$$\therefore L = \underbrace{-\sum \frac{\vec{\omega}^T \cdot \vec{x} + \omega_0}{\|\vec{\omega}\|} \cdot y_i}_{f(x)} \quad \text{s.t.} \quad \underbrace{\|\vec{\omega}\| - 1}_{g(x)} = 0$$

$f(x)$ $g(x)$

$$\therefore L = \underset{\vec{\omega},\lambda}{\operatorname{argmin}} \; -\sum(\vec{\omega}^T\cdot\vec{x}+\omega_0)\cdot y_i + \lambda(\|\vec{\omega}\|-1) = h(\vec{\omega},\lambda)$$

$$\therefore \nabla h = \begin{bmatrix} \frac{\partial}{\partial\vec{\omega}} h(\vec{\omega},\lambda) \\ \frac{\partial}{\partial\lambda} h(\vec{\omega},\lambda) \end{bmatrix} \begin{matrix} \longrightarrow \text{component-1} \\ \longrightarrow \text{component-2} \end{matrix}$$

Before calculating these components, let's see 3 interesting results:

$$\|\vec{\omega}\| = \sqrt{\omega_1^2+\omega_2^2+\dots+\omega_n^2}$$

$$\|\vec{\omega}\| = \sqrt{[\omega_1,\omega_2,\dots\omega_n]\cdot\begin{bmatrix}\omega_1\\ \omega_2\\ \vdots\\ \omega_n\end{bmatrix}}$$

$$\boxed{\|\vec{\omega}\| = \sqrt{\vec{\omega}^T\cdot\vec{\omega}}} \text{ — (I)}$$

$$\frac{d}{dx}\sqrt{x} = \frac{d}{dx}x^{\frac{1}{2}}$$

$$= \frac{1}{2}x^{\frac{1}{2}-1}$$

$$= \frac{1}{2}x^{-\frac{1}{2}}$$

$$= \frac{1}{2x^{1/2}}$$

$$\boxed{\frac{d}{dx}\sqrt{x} = \frac{1}{2\sqrt{x}}} \text{ — (II)}$$

$$\boxed{\frac{\partial}{\partial x}\vec{x}^T\cdot\vec{x} = 2\vec{x}} \text{ — (III)}$$

— Try to prove it yourself.

First component :- $\frac{\partial}{\partial\vec{\omega}} h(\vec{\omega},\lambda)$:

$$\frac{\partial}{\partial\vec{\omega}} h(x,\lambda) = -\sum\frac{\partial}{\partial\vec{\omega}}(\vec{\omega}^T\cdot\vec{x}+\omega_0)\cdot y_i + \frac{\partial}{\partial\vec{\omega}}\lambda(\|\vec{\omega}\|-1)$$

$$= -\sum y_i(\vec{x} + 0) + \lambda\left(\frac{\partial}{\partial \vec{\omega}}\sqrt{\vec{\omega}^T\cdot\vec{\omega}} - 0\right)$$

$$= -\sum y_i \vec{x} + \lambda\left(\frac{1}{2\sqrt{\vec{\omega}^T\cdot\vec{\omega}}} \cdot \frac{\partial}{\partial \vec{\omega}} \vec{\omega}^T\cdot\vec{\omega}\right) \text{ — from (II)}$$

$$= -\sum y_i x_i + \lambda\left(\frac{1}{\cancel{2}\sqrt{\vec{\omega}^T\cdot\vec{\omega}}} \cdot \cancel{2}\vec{\omega}\right) \text{ — from (III)}$$

$$= -\sum y_i x_i + \lambda\left(\frac{\vec{\omega}}{\sqrt{\vec{\omega}^T\cdot\vec{\omega}}}\right)$$

$$\boxed{\frac{\partial}{\partial \vec{\omega}} h = -\sum y_i x_i + \lambda \frac{\vec{\omega}}{\|\vec{\omega}\|}}$$

Second component: $\frac{\partial}{\partial \lambda} h(x, \lambda)$:

$$= \frac{\partial}{\partial \lambda}\left[-\sum(\vec{\omega}^T\cdot\vec{x} + \omega_0)\cdot y_i + \lambda(\|\vec{\omega}\| - 1)\right]$$

$$= -\sum\left(0 + (\|\vec{\omega}\| - 1)\cdot 1\right)$$

$$\boxed{\frac{\partial}{\partial \lambda} h(x, \lambda) = -\sum(\|\vec{\omega}\| - 1)}$$