Revision

<u>objective</u>

find the gradient of the Loss function

using gradient descent $\longrightarrow$ find the values of w_0 $\bar{w}$ where loss function is minimum.

$$\bar{w}_{new} = \bar{w}_{old} - \eta\left(\nabla \bar{w}\right)$$

?

$$w_{0(new)} = w_{0\,old} - \eta\left(\nabla w_0\right)$$

?

Prove-1

$$\nabla \vec{a}^T . \vec{x} = \vec{a}$$

$$\vec{a}^T \cdot \vec{x} = a_1 x_1 + a_2 x_2 + a_3 x_3$$

$$f(x_1, x_2, x_3) = a_1 x_1 + a_2 x_2 + a_3 x_3$$

$$\vec{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \text{ constant}$$

$$\nabla_{\bar{x}} f(\bar{x}) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(\bar{x}) \\ \frac{\partial}{\partial x_2} f(\bar{x}) \\ \frac{\partial}{\partial x_3} f(\bar{x}) \end{bmatrix}$$

$$\frac{\partial}{\partial x_1} f(\bar{x})$$

$$\frac{\partial}{\partial x_1} (a_1 x_1 + a_2 x_2 + a_3 x_3)$$

$$\frac{\partial}{\partial x_1} a_1 x_1 + \frac{\partial}{\partial x_1} a_2 x_2 + \frac{\partial}{\partial x_1} a_3 x_3$$

$$a_1 + 0 + 0$$

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$\nabla_{\bar{x}} f(\bar{x}) = \underbrace{\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}}_{\vec{a}}$$

$$\boxed{\nabla \vec{a}^T \cdot \vec{x} = \vec{a}}$$

Prove-2

$$\nabla \vec{x}^T . \vec{x} = \overrightarrow{2x}$$

$$\vec{x}^T \cdot \vec{x} =$$

$$\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_1^2 + x_2^2 + x_3^2 = f(\bar{x})$$

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$\frac{\partial}{\partial x_1} x_1^2 + x_2^2 + x_3^2$$

$$\underbrace{\frac{\partial}{\partial x_1} x_1^2}_{2x_1} + \underbrace{\frac{\partial}{\partial x_1} x_2^2}_{0} + \underbrace{\frac{\partial}{\partial x_1} x_3^2}_{0}$$

$$\nabla_{\bar{x}} f(\bar{x}) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(\bar{x}) \\ \frac{\partial}{\partial x_2} f(\bar{x}) \\ \frac{\partial}{\partial x_3} f(\bar{x}) \end{bmatrix}$$

$$\nabla_{\bar{x}} f(\bar{x}) = \begin{bmatrix} 2x_1 \\ 2x_2 \\ 2x_3 \end{bmatrix} = 2 \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$\boxed{\nabla \bar{x}^T \cdot \bar{x} = 2\bar{x}}$$

find that value of $\bar{w}$, w_0 where the Loss function is min.

Optimisation Problem

function

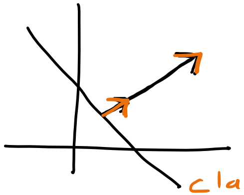
$$\min_{\bar{w}, w_0} - \sum_i \left(\frac{\vec{w}^T . \vec{x} + w_0}{\|\vec{w}\|} \right) . y_i$$

$$\|\bar{w}\| = \sqrt{w_1^2 + w_2^2 + \dots + w_n^2}$$

$\|\bar{w}\| = 1$ (constraint)

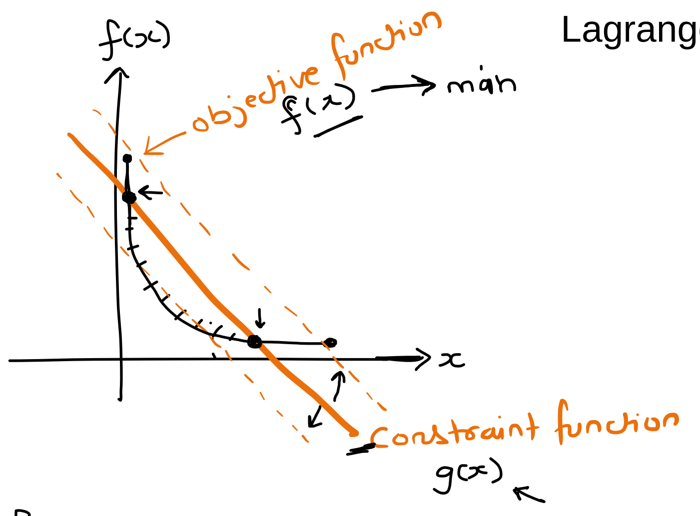
True label

x_1	x_2	y
		-1
		$+1$
		-1
		-1



classifier

Lagrange Multipliers



$\lambda \rightarrow \uparrow \downarrow$

factory Boots

for how much time I should run the machine to get max. prod

machine

not run

$\boxed{2 < x < 8}$

$f(x) \longrightarrow$ objective

$g(x) \longrightarrow$ constraint

$f(\bar{x}) = \boxed{x^2 + y^2}$

$g(\bar{x}) = 2x - 1 = 0$

$2x = 1$

$$\left.\begin{array}{l} \min\limits_{\bar{x}} \; f(\bar{x}) \\ \underline{ST} : \; g(\bar{x}) = 0 \end{array}\right\} \longrightarrow \min_{\bar{x}} \; \overbrace{\mathcal{L} \; f(x) + \lambda (g(x))}^{\text{Lagrange function}}$$

$0 > 0$

$$f(x) = x^2 - 3x - 3$$

$$ST: \; -x^2 + 2x + 3 = 0$$

Example-1

find min of $f(x) = x^2 - 3x - 3 \longrightarrow$

S.T $\quad -x^2 + 2x + 3 \longrightarrow g(x)$

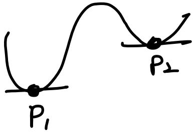
$\rightarrow f(x) + \lambda \; g(x)$

$$\min \mathcal{L}(x, \lambda) = x^2 - 3x - 3 + \lambda(-x^2 + 2x + 3)$$

$$\bar{z} = \begin{bmatrix} x \\ \lambda \end{bmatrix}$$

$$\min_{\bar{z}} \mathcal{L}(\bar{z}) = x^2 - 3x - 3 + \lambda(-x^2 + 2x + 3)$$

$$\nabla_{\bar{\omega}} \mathcal{L}(\bar{z}) = \begin{bmatrix} \frac{\partial}{\partial x} \mathcal{L}(\bar{z}) \\ \frac{\partial}{\partial \lambda} \mathcal{L}(\bar{z}) \end{bmatrix} = \begin{bmatrix} 2x - 3 - 0 - 2\lambda x + 2\lambda + 0 \\ 0 - 0 - 0 \; -x^2 + 2x + 3 \end{bmatrix} \begin{bmatrix} 2x - 3 - 2\lambda x + 2\lambda \\ -x^2 + 2x + 3 \end{bmatrix}$$



Minimum will occur at the point where gradient of the function is zero

$$\begin{bmatrix} \frac{\partial}{\partial x} \mathcal{L}(\bar{z}) \\ \frac{\partial}{\partial \lambda} \mathcal{L}(\bar{z}) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Example-1

$$2x - 3 - 2\lambda x + 2\lambda = 0$$

$$2x - 3 = 2\lambda x - 2\lambda$$

$$\left[\frac{2x-3}{2(x-1)} = \lambda\right]$$

$$\lambda = \frac{2\times 3 - 3}{2(3-1)} = \frac{3}{4}$$

$$\lambda = \frac{2\times -1 - 3}{2(-1-1)} = \frac{5}{4}$$

$$\lambda = \left(\frac{3}{4}, \frac{5}{4}\right) \quad \text{min at } x = 3 \ \& \ -1$$

at ($x = 3$ and $\lambda = 3/4$) output $\mathcal{L}(z) = -3$

at ($x = -1$ and $\lambda = 5/4$) output $\mathcal{L}(z) = 1$

(x, y)

$(3, -3)$

$(-1, 1)$

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

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

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

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

$$[x = 3, -1]$$

At $x = 3$ & -1
my function is minimum

$$\mathcal{L}(z) = x^2 - 3x - 3 + \lambda(-x^2 + 2x + 3)$$

$$= 9 - 9 - 3 + \frac{3}{4}(-9 + 6 + 3)$$

$$= -3$$

$$\mathcal{L}(z) = 1 + 3 - 3 + \lambda(-1 - 2 + 3)$$

$$= 1 + \frac{5}{4}(0)$$

$$= 1$$

Solving Loss Function Using Lagrange Multipliers

$g(x) = 0$

Loss function

$$||\vec{\omega}|| = \sqrt{\bar{\omega}^T \cdot \bar{\omega}}$$
$$= \sqrt{\omega_1^2 + \omega_2^2 + \cdots \omega_n^2}$$

$$\min_{\bar{\omega}\, \omega_0} - \sum_i \left(\frac{\vec{w}^T . \vec{x} + w_o}{||\vec{w}||} \right) . y_i \rightarrow f(x)$$

$$ST: ||\vec{w}|| = 1 \rightarrow g(x)$$

$$\boxed{||\bar{\omega}|| - 1 = 0}$$

$$\min_{\bar{\omega}, \omega_0, \lambda} - \sum_i (\bar{\omega}^T \cdot \bar{x} + \omega_0) \cdot y_i + \lambda \left[\sqrt{\bar{\omega}^T \cdot \bar{\omega}} - 1 \right]$$

$$- \sum_i \overbrace{(\bar{\omega}^T \cdot \bar{x} \cdot y_i + \omega_0 y_i)}^{1} + \overbrace{\lambda \left[\sqrt{\bar{\omega}^T \cdot \bar{\omega}} - 1 \right]}^{2}$$

$$\nabla_{\bar{\omega}} \mathcal{L}(\bar{\omega}, \omega_0, \lambda)$$

$$\frac{\partial}{\partial \bar{\omega}} \bar{\omega}^T \cdot \bar{x} \cdot y_i + \frac{\partial}{\partial \bar{\omega}} \omega_0 y_i$$

$$[\bar{x} \cdot y_i] \quad 1$$

Solving Loss Function Using Lagrange Multipliers

$$\lambda \left[\sqrt{\bar{\omega}^T \cdot \bar{\omega}} - 1 \right]$$

$$\lambda \sqrt{\bar{\omega}^T \cdot \bar{\omega}} - \lambda$$

$$\frac{\partial}{\partial \bar{\omega}} \lambda \sqrt{\bar{\omega}^T \cdot \bar{\omega}} - \underbrace{\frac{\partial}{\partial \bar{\omega}} \lambda}_{0}$$

$$\lambda \frac{\partial}{\partial \bar{\omega}} \sqrt{\bar{\omega}^T \cdot \bar{\omega}}$$

$$\lambda \frac{1}{2\sqrt{\bar{\omega}^T \cdot \bar{\omega}}} \cdot \frac{\partial}{\partial \bar{\omega}} \bar{\omega}^T \cdot \bar{\omega}$$

$$\lambda \frac{1}{\cancel{2}\sqrt{\bar{\omega}^T \cdot \bar{\omega}}} \cdot \cancel{2}\bar{\omega} = \lambda \frac{\bar{\omega}}{\|\bar{\omega}\|} \quad 2$$

$$\sqrt{x}$$

$$(x)^{1/2}$$

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

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

$$\frac{1}{2\sqrt{x}}$$

Solving Loss Function Using Lagrange Multipliers

$$\boxed{\nabla_{\bar{\omega}} \mathcal{L}(\bar{\omega}, \omega_0, \lambda) = -\sum_{i=1}^{n} \bar{x} \cdot y_i + \lambda \frac{\bar{\omega}}{\|\omega\|}}$$

Gradient with respect to $\bar{\omega}$

$$\mathcal{L}(\bar{\omega}, \omega_0, \lambda) = -\sum_{i=1}^{n} \overbrace{(\bar{\omega}^T \cdot \bar{x} + \omega_0) \cdot y_i}^{1} + \lambda \overbrace{\left(\sqrt{\bar{\omega}^T \cdot \bar{\omega}} - 1\right)}^{2}$$

$$\nabla_{\omega_0} \mathcal{L}(\bar{\omega}, \omega_0, \lambda)$$

$$= -\underbrace{\frac{\partial}{\partial \omega_0} \bar{\omega}^T \cdot \bar{x} \cdot y_i}_{0} - \underbrace{\frac{\partial}{\partial \omega_0} \omega_0 \cdot y_i}_{y_i} + \underbrace{\frac{\partial}{\partial \omega_0} \lambda \sqrt{\bar{\omega}^T \cdot \bar{\omega}}}_{0} - \underbrace{\frac{\partial}{\partial \omega_0} \lambda}_{0}$$

$$\nabla_{\omega_0} \mathcal{L}(\bar{\omega}, \omega_0, \lambda) = -\sum_{i=1}^{n} y_i \quad \text{or} \quad \sum_{i=1}^{n} (-y_i)$$

Sumit.shukla-1@g
scaler.com

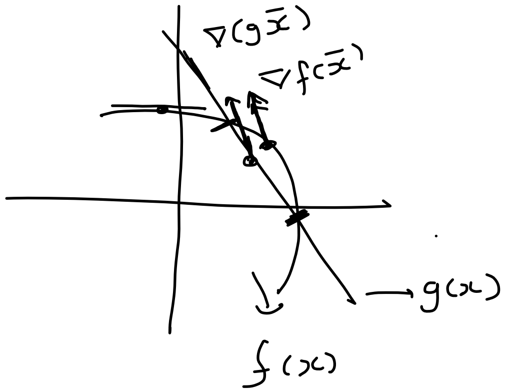
Solving Loss Function Using Lagrange Multipliers

Summary

1. we wanted to build a classifier
2. we obtained a loss function $-\sum_{i=1}^{n}\left(\frac{\bar{w}^T\bar{x}+w_0}{||w||}\right)\cdot y_i$
3. find $\bar{w}, w_0$ such that loss function is minimum
4. we decided to use gradient descent $\left[\begin{array}{l} \bar{w}^{(new)} = \bar{w}^{(old)} - \eta\,(\nabla\bar{w}) \\ w_0^{(new)} = w_0^{(old)} - \eta\,(\nabla w_0) \end{array}\right]$
5. It was not easy to find gradient of the function (Loss) w.r.t $\bar{w}, w_0$
6. Assumption $||\bar{w}|| = 1$
7. New Loss function (Lagrange function) $\left[\mathcal{L}(\bar{w}, w_0, \lambda) = -\sum_{i}(\bar{w}^T\cdot x + w_0)y_i + \lambda\left[||\bar{w}|| - 1\right]\right]$
8. $\nabla_{\bar{w}}\mathcal{L}(\bar{w}, w_0, \lambda) = -\sum_{i=1}^{n}\bar{x}\cdot y_i + \lambda\left(\frac{\bar{w}}{||\bar{w}||}\right)$

   $\nabla_{w_0}\mathcal{L}(\bar{w}, w_0, \lambda) = -\sum_{i=1}^{n} y_i$

Solving Loss Function Using Lagrange Multipliers



$$\nabla f(\bar{x}) = \lambda \nabla g(\bar{x})$$