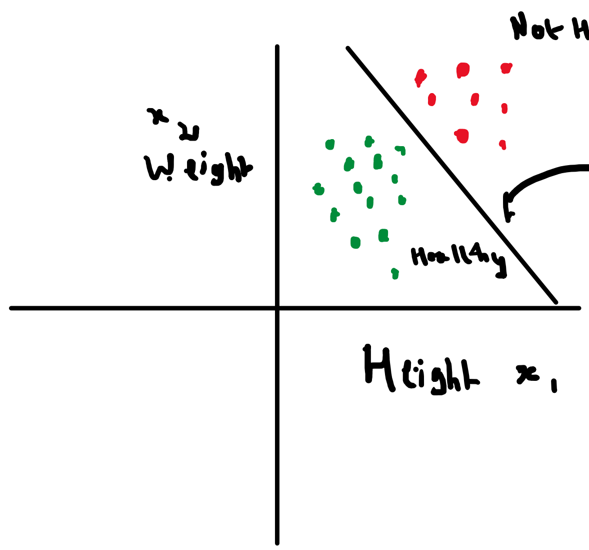
Gradient Descent Recap

10 November 2025 14:10

ML Goal => Get right boundary / right equation.

What will minimizing it do? [Minimize Loss]

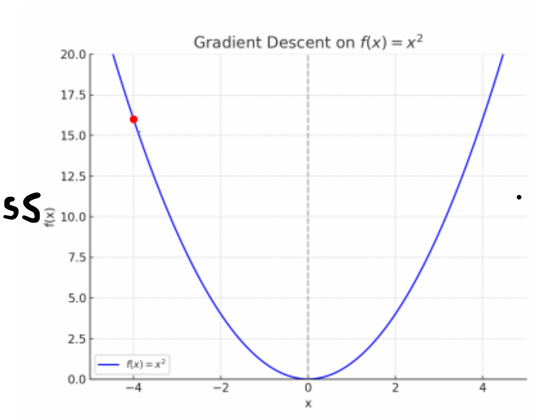


$w^T x + w_0 = 0$

G.D which is part of convex optimization

How do we minimize loss fn?

Loss



for simplicity, taking

$L = w_1^2$

(Gradient → Maxima)

+

$$W_{new} = W_{old} - \lambda \frac{dL}{dW}$$

[λ = 0.01
0.5
0.1]

New   old − ⊤ dW   [0.1]

stepsize

G.D guarantee global Minima ?

Convex fns → G.D guarantee

[Non Convex fns → no guarantee] ?

[Multiple local Minimas.

Weight can get stuck at a local Minimum]

What all do we need ?

i) Weights ⇒ random weight.
ii) λ - learning rate
iii) Iterations / epochs.
iv) Loss function
v) Define a function for derivative

G.D ⇒ derivative.

Unconstrained Optimization

04 September 2025 18:50

How can I minimize a function $[3x^2 + 4$?$]$

$$3 \times 0^2 + 4 = 4$$

$$\frac{d(3x^2+4)}{dx} = 6x$$

(No constraints or restrictions)

$$6x = 0$$

$[x = 0]$ optimal x

Minimum value $= 4$

Constrained Optimization

04 September 2025 18:52

$\left[y = x^2 \qquad y = x + 3 \right]$ [intersecting points]

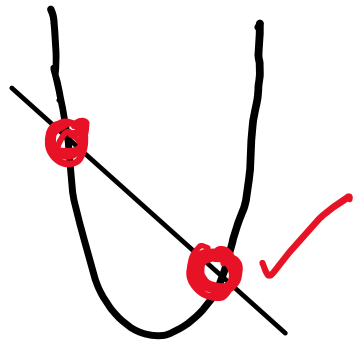
Minimize $f(x) = x^2$ such that

$[y = x + 3]$

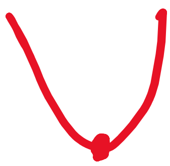
$x^2 = x + 3$

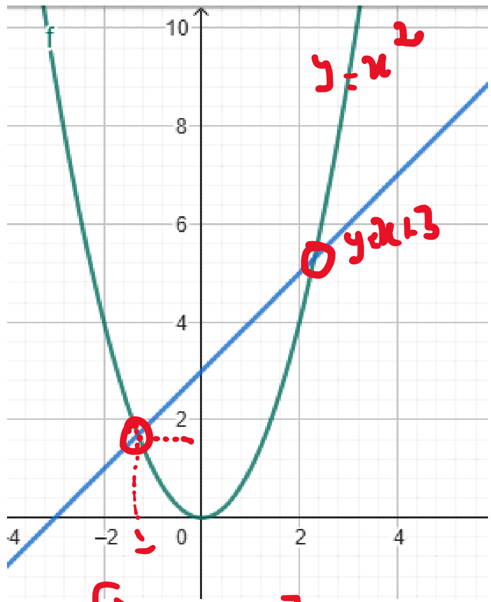
$[x^2 - x - 3 = 0] \Rightarrow \left(\frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \right)$

2 solutions $\Rightarrow$ intersecting points



Min x^2





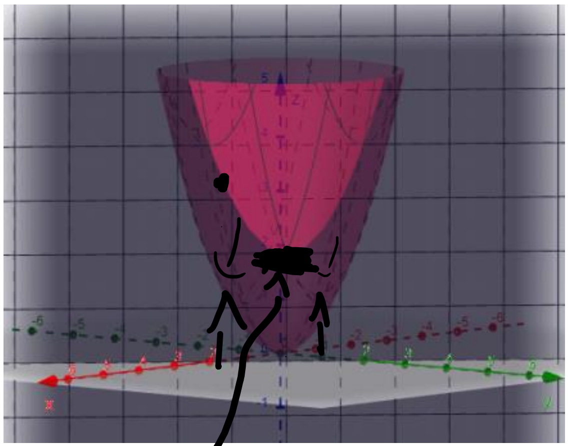
$$\begin{bmatrix} x = -1.8 \\ y = 1.9 \end{bmatrix}$$

Constrained optimization in multiple dimensions

04 September 2025 19:29

Minimize $(x^2 + y^2)$

$(x + y - 2 = 0)$



$[\phi(x,y): x^2 + y^2 \Rightarrow]$ Minimize

$g(x) \rightarrow \underbrace{x + y - 2 = 0}_{\text{Constraint}}$

$(1, 1)$

Value of fn $= 2$

Minimize $[\phi(x) (\pm) \underbrace{\lambda}_{\text{Lagrangian Multiplier}} g(x,y)]$ $\lambda = 0$

2

$h(x, y, \lambda)$

$1 \quad 1 + \quad \cancel{-2}\,(\cancel{1+1-2})$

Minimize $((x^2 + y^2 + \lambda(x + y - 2)))$

Minimize $(\underbrace{x^2 + y^2 + \lambda x + \lambda y - 2\lambda}_{})$

$0 \quad 0 \qquad x \quad + y \quad -2$

$\frac{\partial h}{\partial x} \Rightarrow 2x + \lambda = 0$ (1)

$2x = -\lambda$

$x = -\lambda/2 \Rightarrow -\frac{(-2)}{2} = 1$

$\frac{\partial h}{\partial y} \Rightarrow 2y + \lambda = 0$ (2)

$y = -\lambda/2 \quad -\frac{(-2)}{2} = 1$

$\frac{\partial h}{\partial \lambda} \Rightarrow x + y - 2 = 0$ (3)

$\frac{\delta h}{\delta \lambda} \Rightarrow \quad x + y - 2 = 0 \quad (3)$

$-\lambda/2 - \lambda/2 - 2 = 0$

$-\lambda - 2 = 0$

$-\lambda = 2$

$[\lambda = -2]$

x, y, λ

$x^2 + y^2 - \lambda(x + y - 2)$

$1^2 + 1^2 - 2(1 + 1 - 2)$

$\left(1 + 1 - 2 \times 0 = 2 \right)$

$$\left[\begin{array}{c} \text{Minimum value is } \quad z = 2 \\ \text{or} \\ f(x,y) = 2 \end{array} \right] \text{ where } x,y = (1,1)$$

$$\left(\begin{array}{c} \text{Lagrangian multipliers ?} \\ \\ \text{Where ?} \end{array} \right)$$

Largrangian Recap

04 September 2025 20:09

"Lagranges multipliers".

Constraint optimisation problem $\longrightarrow$ $\boxed{\text{Lagranges Multiplier}}$ $\longrightarrow$ UnConstraint optimisation problem

$$\min_{\bar{x}} f(\bar{x}) \quad \text{s.t.} \quad g(\bar{x}) = 0$$

$$\min_{\bar{x}} \underbrace{f(\bar{x}) \pm \lambda(g(\bar{x}))}$$

[Multiple dimensions + Multiple constraints]

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

$$\hookrightarrow \quad f(\underbrace{x, y, z}) + \lambda_1 (g(x, y, z)) + \lambda_2 g_2(x, y, z) + \lambda_3 g_3(x, y, z) + \lambda_4 g_4(x, y, z)$$

Our loss function

04 September 2025 20:04

Direct

Neg class $\ominus$ $+$ Pos class

$$= \min_{\bar{\omega}, \omega_0} -\left(\sum_{i=1}^{n} \left(\frac{\bar{\omega}^T \bar{x} + \omega_0}{\|\bar{\omega}\|} \right) y_i \right)$$

$$-\sum (w^T x + w_0) y_i$$

$$\sqrt{w_1^2 + w_2^2} = 1$$

$\|\bar{w}\|$ => Magnitude of my weight vector

Deep Learning problems => [billions]

$$\left[\sqrt{w_1^2 + w_2^2 + w_3^2 \cdots \cdots w_{1000000}^2} \right]$$

$$\left[\min -\sum (w^T x + w_0) y_i \right.$$

wrt constraint $\left. \right]$

$$\sqrt{w_1^2 + w_2^2} = 1$$

$$\boxed{\|w\|^2 = 1}$$

$$\min -\sum \left[(w^T x + w_0) y_i + \lambda (\|w\|^2 - 1) \right]$$

$$\left[\frac{\partial \mathcal{L}}{\partial w_0} = -\sum_{i=1}^{n} y_i \right] \quad \Sigma$$

$$\left[\frac{\partial \mathcal{L}}{\partial \lambda} = \|\mathbf{w}\|^2 - 1 \right]$$

$$\left[\frac{\partial L}{\partial w}, \frac{\partial L}{\partial w_0}, \frac{\partial L}{\partial \lambda} \right]$$

Update the weights.

Primal update (gradient descent on L):

$$\left[\begin{aligned} \mathbf{w}_{t+1} &= \mathbf{w}_t - \alpha \nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}_t, \lambda_t) \\ w_{0,t+1} &= w_{0,t} - \alpha \frac{\partial \mathcal{L}}{\partial w_0} \end{aligned} \right]$$

[Dual update (gradient ascent on L):]

$$\left[\lambda_{t+1} = \lambda_t + \beta \frac{\partial \mathcal{L}}{\partial \lambda} \right]$$

[Constraint]!!

[constraint]!!

Primal Dual Gradient Descent

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \alpha\left(-\sum_i y_i \mathbf{x}_i + 2\lambda_t \mathbf{w}_t\right) = \mathbf{w}_t + \alpha\sum_i y_i\mathbf{x}_i - 2\alpha\lambda_t\mathbf{w}_t,$$

$$w_{0,t+1} = w_{0,t} + \alpha\sum_i y_i,$$

$$\lambda_{t+1} = \lambda_t + \beta\left(\|\mathbf{w}_t\|^2 - 1\right).$$

Lagrangian multipliers ⇒ Make non-convex loss Fns easier to solve

Gain fn ⇒ Max $\sum d_1 + d_2 + d_3 + d_4 \ldots\ldots$

Loss Fn: $- \sum \frac{(w^T x + w_0)}{\|w\|} y_i$

distance	ypred
−6	−1
−1	−1
3	1
7	1
8	1

$\frac{-7}{20}$ $\frac{10}{20}$ ⇒ +ve

$\frac{-7}{2000}$ $\frac{10}{\ldots}$ ⇒ +ve

$$\frac{-7}{2000}$$

$$\frac{10}{2} \Rightarrow +ve$$

$$-\sum \frac{(w^T x + w_0)\, y_i}{\|w\|}$$

Norm

$\|w\|$

negative

$$\frac{10}{2000} \Rightarrow +ve$$

Finance Example

16 September 2025 20:33

60% 40%

	PrevDayReturn (w_1 80%)	PrevWeekReturn (w_2 20%)	Target	
S1	0.01	0.03	BUY	+1
S2	-0.02	-0.05	SELL	-1
S3	0.015	0.01	BUY	+1
S4	-0.005	-0.01	SELL	-1
S5	0.02	0.025	BUY	+1
S6	-0.015	-0.02	SELL	-1

equation

I want weights such that I can analyse wrt

$\boxed{\%}$

[Loss fn]

$[w_1 + w_2 = 1]$

[Lagrangian Multipliers]

ML Use cases of Lagrangian

10 September 2025 19:26

i) PCA => Principal Component Analysis

ii) [SVM => Support Vector Machine]

iii) Simplify non-convex loss functions

Apache Helicopter

16 September 2025 17:27

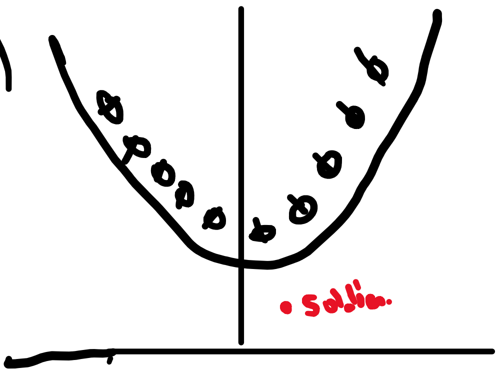
Apache helicopter

An Apache helicopter of the enemy is flying along the curve given by $y = x^2 + 7$.

A soldier, placed at $(3, 7)$, wants to shoot down the helicopter when it is nearest to him.

What is the nearest distance between the soldier and the helicopter ?

(Minimize)



at any given point the position of the helicopter can be represented by (x, y)

(x, x^2+7)

$(3, 7)$

(Distance) Euclidean dist?

$$d = \sqrt{(3-x)^2 + (7-(x^2+7))^2}$$

$\dfrac{\text{Min } d}{x}$ $\qquad \dfrac{dd}{dx}$ $\qquad \sqrt{a}\Downarrow$

$(\sqrt{a})\Downarrow$

$$d = \left(\sqrt{\left[(3-x)^2 + (7-(x^2+7))^2\right]}\right)$$

$$(3-x)^2 + (x^2)^2$$

$$9 + x^2 - 6x + x^4$$

$$9 + x^2 - 6x + x^4$$

-1

$$\left[2x - 6 + 4x^3 = 0\right]$$

2×-1

$-2 - 1$

$$4x^3 + 2x - 6 = 0$$

$x = 1$

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

$2 + 1 - 3 = 0$

$$\left[x = \underline{1} \text{ or } -\underline{1}\right]$$