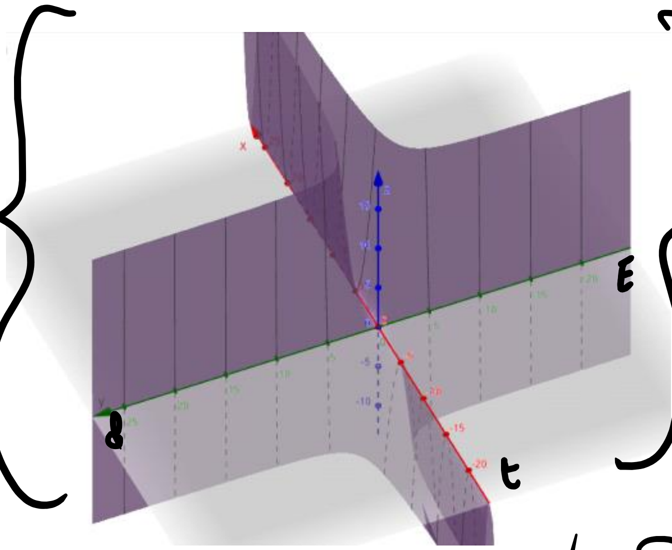
Partial Derivatives

05 November 2025 14:12

3D figure



d = distance

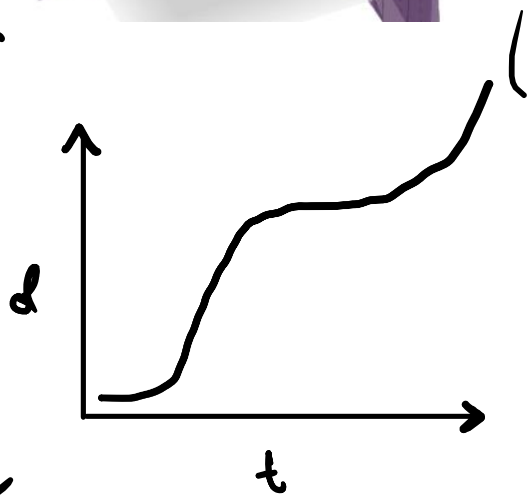
E = Experience of driver

t = time

$$d(E,t) = \boxed{Et^2}$$

$$\frac{\partial d}{\partial E} = t^2 \qquad \{t^2\}$$

$$\frac{\partial d}{\partial t} = 2Et$$



$$d(t) = 3t^3 - 3t^2 + 2t \qquad \frac{\partial d}{\partial t}$$

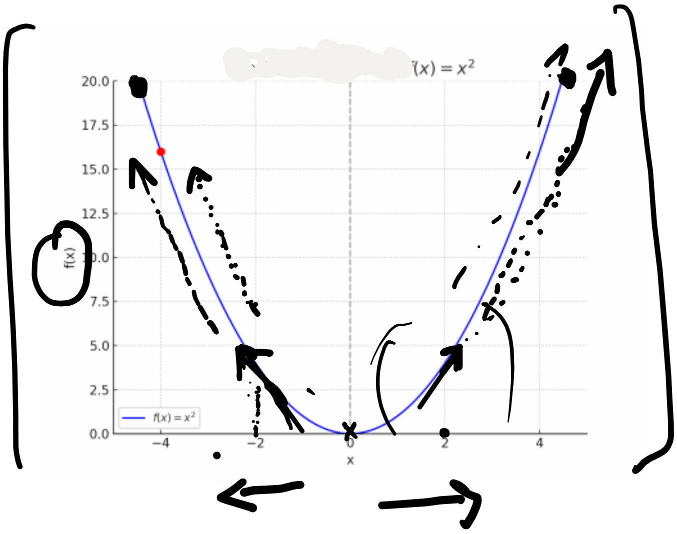
$\Rightarrow$ 1st derivative

Slope at a given point

Direction of slope

05 November 2025 20:57

$x = 2$



-4

$y = x^2$

$\frac{dy}{dx} = 2x$ $\quad x = 2$

4

x is negative

$x = -2$

$\frac{dy}{dx} = 2x = 2 \times -2 = -4$

Maximum value of the function

x is positive

$x = +2$

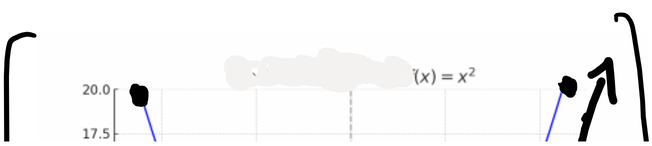
$\frac{dy}{dx} = 2 \times 2 = \boxed{4}$

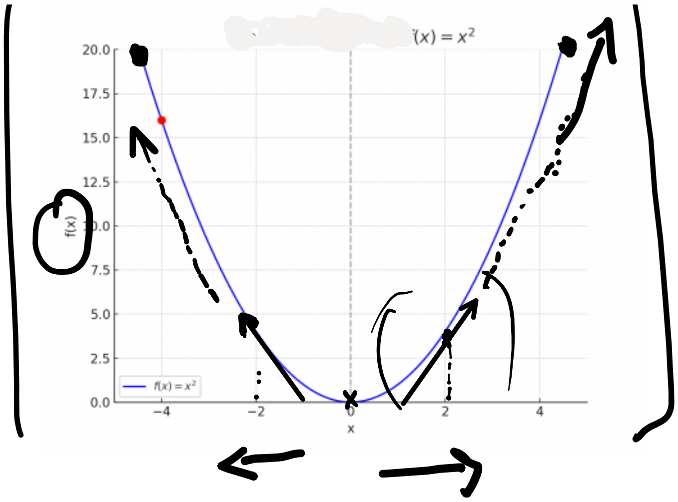
[derivative always points towards Max value of the function]

Max is 0

$y = -x^2$

$-2x$





$-2x$

Negative value of x

-2

4

Gradient

05 November 2025 14:23

$$\boxed{d = Et^2}$$

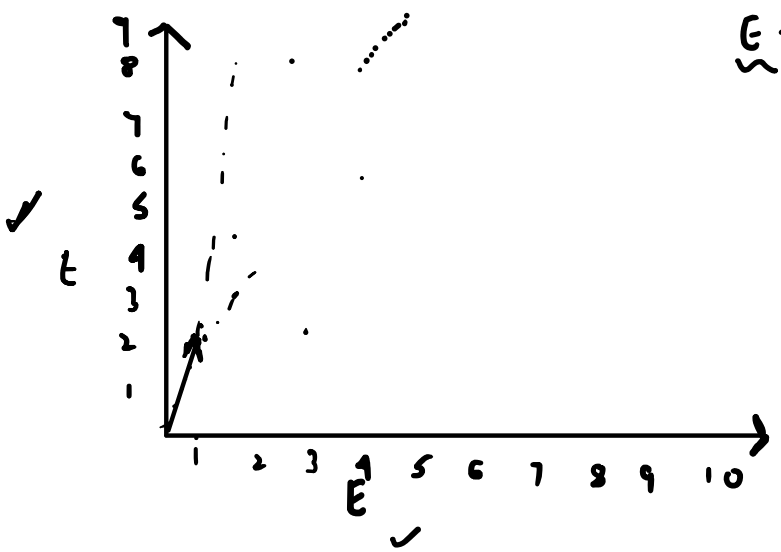
$$\frac{dd}{dE} = t^2 \qquad \frac{dd}{dt} = 2Et$$

$$\left[t^2,\ 2Et\right]$$ Gradient

[Vector]

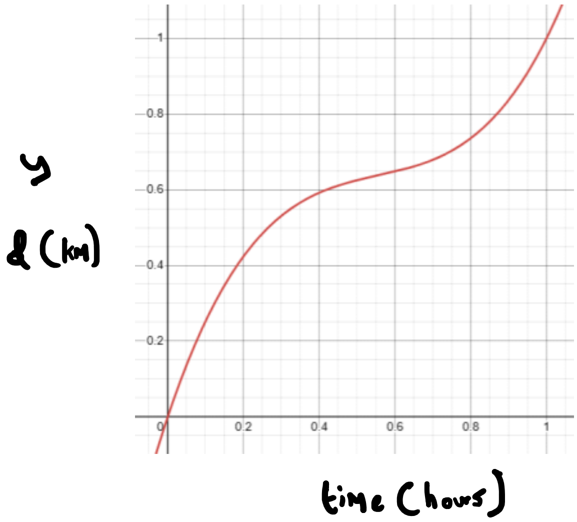
$t = 1$, $E = 1$

$$\left[1^2,\ 2 \times 1 \times 1\right] = \left(\left[1, 2\right]\right)$$



Gradient always points towards the MAXIMUM value of the fn

1 variable → 1 number



$d = 3t^3 - 5t^2 + 3t$

What is a gradient?

{ Vector ✡ Maximum ✡ of function }

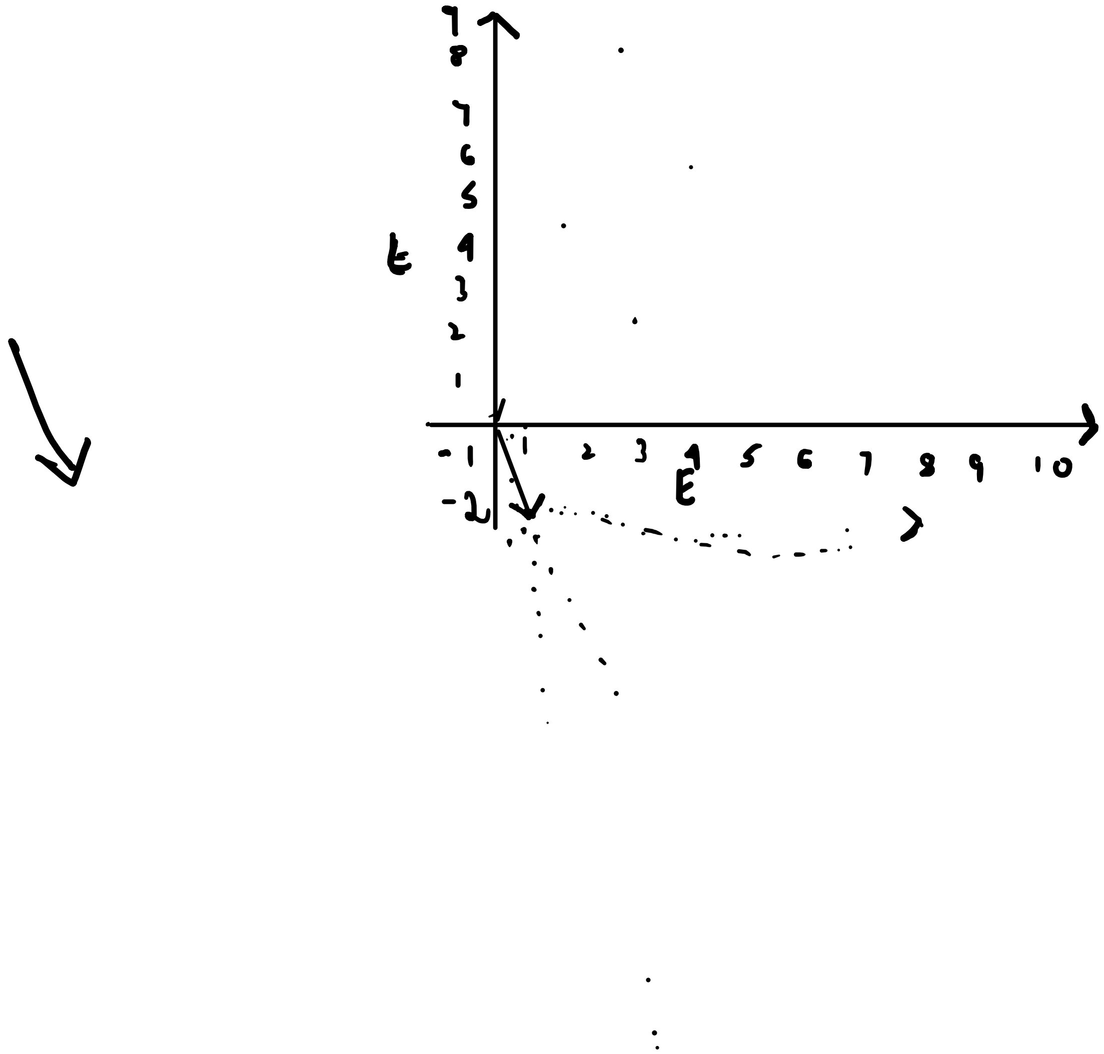
[Core of Convex optimization ✡]

$t = 1$ ✓

$E = -1$ ✓

$[t^2, 2Et]$ ✓ $\quad 1, \; 2 \times -1 \times 1$

$\boxed{\text{Vector}}$ $\quad 1, -2$


t
E

Convex Optimization Techniques

05 November 2025 14:41

Set of optimization techniques, that help us find the minima of a given function

i) Gradient Descent ✓ First derivative

ii) Newtons' Methods

iii) Quasi Newtons Methods

Hessian

iv) Lagrangian Multipliers ✓

Gradient Descent

05 November 2025 14:40

$$\text{Gain} = \sum \frac{(w^t x + w_0)\, y_i}{\|w\|} \quad \text{Maximize}$$

$$\text{Loss} = \left(- \sum \frac{(w^t x + w_0)\, y_i}{\|w\|} \right) \quad \text{Minimize}$$

$$[W^t x + W_0]$$

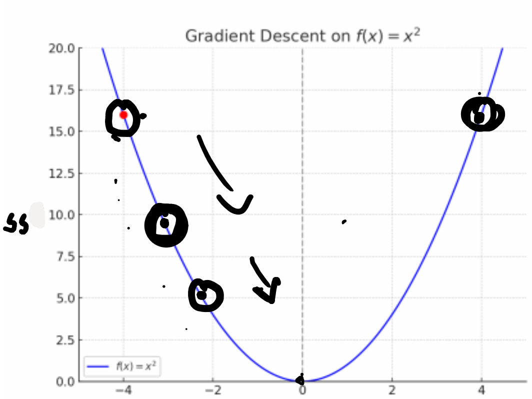
$$[\text{Loss} = w^2]$$

Why do we minimize loss fn?

optimal weights => equation

↳ Boundary

Loss



w

[Boundary / Weights.]

Random weight

$w = -4$

Weight that minimizes loss

$l = w^2$

$2w$

$$\left[+ \frac{dl}{dw} \Rightarrow \text{Slope} \right]$$

$$+ \frac{dl}{dw}$$

$$W_{new} = w_{old} - \frac{dl}{}$$

δw

-4

$$\frac{dl}{\delta w} = 2w$$

$-4 - (2 \times -4)$

$-4 - (-8)$

$-4 + 8 = 4$

$w_{new} = 4$

$$W_{new} = W - (2w)$$

$$W_{new} = -4 - (2 \times -4)$$

$$= -4 - (-8)$$

$$= -4 + 8 = 4$$

How much to move

$\nearrow$ Learning rate

$$W_{new} = w_{old} - \lambda \frac{dl}{\delta w}$$

$$\lambda = 0.1$$

$$W_{new} = -4 - 0.1(2 \times -4)$$

$$= -4 - 0.1(-8)$$

$$= -4 + 0.8 = -3.2$$

$= -4 + 0.8 = -3.2$

$W_{old} = -3.2$

$2w = 2 \times -3.2$

$W_{new} = -3.2 - 0.1(2 \times -3.2)$

$= -3.2 - 0.1(-6.4)$

$= -3.2 + 0.64$

$= -2.56$

Gradient Desent

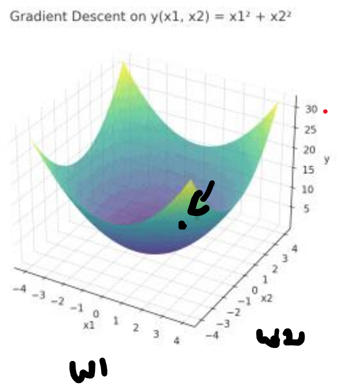
$\left(W_{old} + \frac{\partial l}{\partial w} \right)$ [Gradient Ascent]

Maxima

[Gradient Descent]

Gradient Descent in case of 2 weights

02 September 2025 20:21



0 [0, 0]

$\left[Loss = w_1^2 + w_2^2 \right]$

$([0, 0])$

$(3, -2)$

[Gradient] Descent

$$f(x_1, x_2) = x_1^2 + x_2^2$$

This is a simple convex "bowl"-shaped function with a **global minimum at** $(0, 0)$.

---

1. Define Gradient

The gradient is the vector of partial derivatives:

$$\nabla f(x_1, x_2) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \right) = (2x_1, 2x_2)$$

Let's start with:

$$x^{(0)} = (3, -2), \quad \eta = 0.1$$

$$W = [3, -2]$$

$$Loss = w_1^2 + w_2^2$$

$$[2w_1, 2w_2]$$

$$\eta = 0.1$$

$$W_{new} = [3, -2] - 0.1\,[2 \times 3, 2 \times -2]$$

$$\begin{bmatrix} 3 \\ -2 \end{bmatrix} - 0.1 \begin{bmatrix} 6 \\ -4 \end{bmatrix}$$

$$x^{(0)} = (3, -2)$$

- Gradient at $x^{(0)}$:

$$\nabla f(3, -2) = (2 \cdot 3,\ 2 \cdot (-2)) = (6, -4)$$

$$x^{(0)} = (3, -2)$$

- Gradient at $x^{(0)}$:

$$\nabla f(3, -2) = (2 \cdot 3,\ 2 \cdot (-2)) = (6, -4)$$

- Update:

$$x^{(1)} = (3, -2) - 0.1 \cdot (6, -4)$$

$$x^{(1)} = (3 - 0.6,\ -2 - (-0.4)) = (2.4,\ -1.6)$$

- Current point:

$$x^{(1)} = (2.4, -1.6)$$

- Gradient at $x^{(1)}$:

$$\nabla f(2.4, -1.6) = (2 \cdot 2.4,\ 2 \cdot (-1.6)) = (4.8, -3.2)$$

- Update:

$$x^{(2)} = (2.4, -1.6) - 0.1 \cdot (4.8, -3.2)$$

$$x^{(2)} = (2.4 - 0.48,\ -1.6 - (-0.32)) = (1.92,\ -1.28)$$

Step k	x_1	x_2	$f(x_1, x_2)$
0 (start)	3.00	-2.00	13.00
1	2.40	-1.60	8.32
2	1.92	-1.28	5.3248

$$\begin{bmatrix} 3 \\ -2 \end{bmatrix} \quad \begin{bmatrix} -0.6 \\ +0.4 \end{bmatrix}$$

$$\begin{bmatrix} 2.4 \\ -1.6 \end{bmatrix}$$

$(3, -2)$

$(2.4, -1.6)$

When to stop

02 September 2025 20:21

Derivative λ × Derivative.

	iter	W_{old}	Loss	grad f(x_k)	Δx_k = -η·grad	W_{new}
1	0	-4.0	16.0	-8.0	1.6	-2.4
2	1	-2.4	5.76	-4.8	0.96	-1.44
3	2	-1.44	2.0736	-2.88	0.576	-0.864
4	3	-0.864	0.7464959999999999	-1.728	0.3456	-0.5184
5	4	-0.5184	0.26873855999999996	-1.0368	0.20736	-0.31104
6	5	-0.31104	0.09674588159999999	-0.62208	0.124416	-0.18662399999999998
7	6	-0.18662399999999998	0.03482851737599999	-0.37324799999999997	0.0746496	-0.11197439999999999
8	7	-0.11197439999999999	0.012538266255359998	-0.22394879999999998	0.04478976	-0.06718463999999999
9	8	-0.06718463999999999	0.004513775851929598	-0.13436927999999998	0.026873855999999998	-0.040310783999999999
10	9	-0.040310783999999999	0.0016249593066946552	-0.080621567999999998	0.016124313599999995	-0.024186470399999993
11	10	-0.024186470399999993	0.0005849853504100758	-0.048372940799999986	0.009674588159999997	-0.014511882239999996
12	11	-0.014511882239999996	0.0002105947261476273	-0.029023764479999999	0.005804752895999999	-0.008707129343999996
13	12	-0.008707129343999996	7.58141014131458e-05	-0.017414258687999992	0.003482851737599999	-0.005224277606399997
14	13	-0.005224277606399997	2.7293076508732487e-05	-0.010448555212799995	0.002089711042559999	-0.0031345665638399982
15	14	-0.0031345665638399982	9.825507543143694e-06	-0.006269133127679964	0.0012538266255359993	-0.001880739938303999

) → 0.00002

) → 0.00001

0.

λ = 0.01

λ = 0.1

G.D will stop, when you stop seeing a significant improvement

[Epoch]

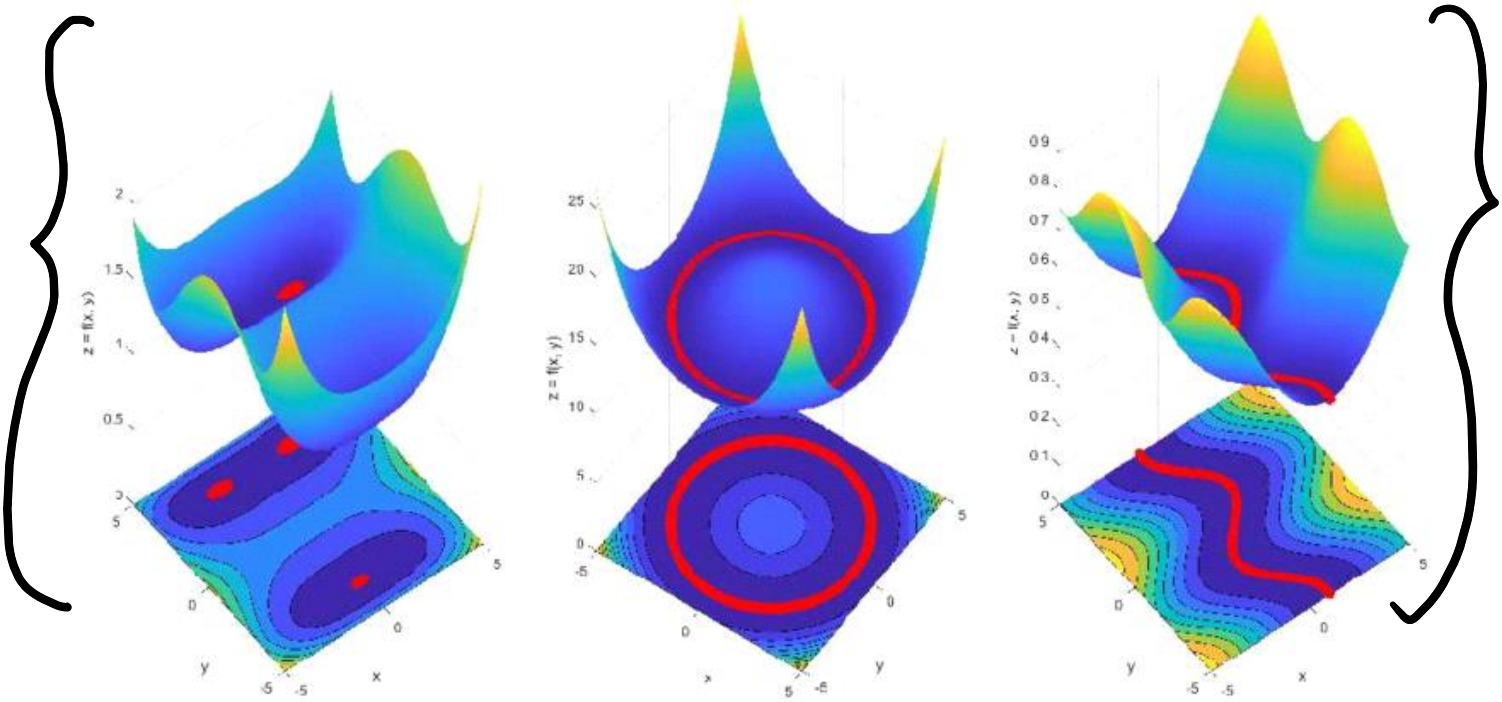
100

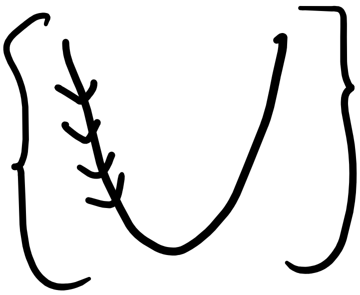
0.0001

loss_decre
asing_and...

Different Shapes of loss

05 November 2025 15:01





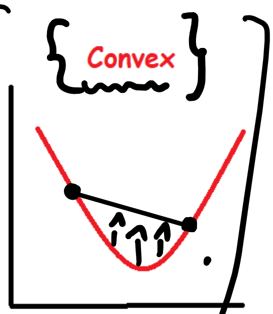
No!!

Convex vs Non Convex funcions

08 September 2025 19:12

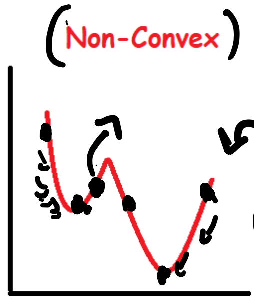
Does Gradient Descent assure a perfect solution? → No!

Minima ?



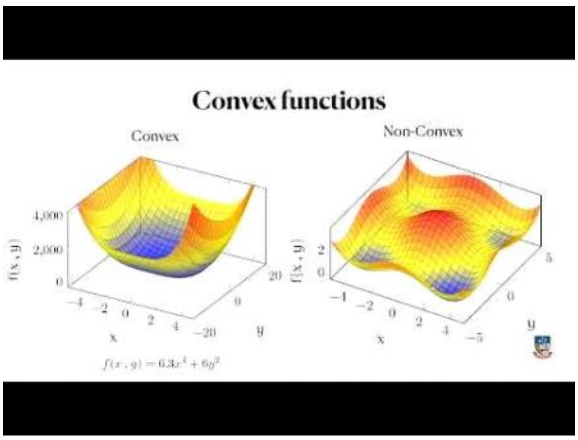
Above the function

→ local Minima and global Minima same

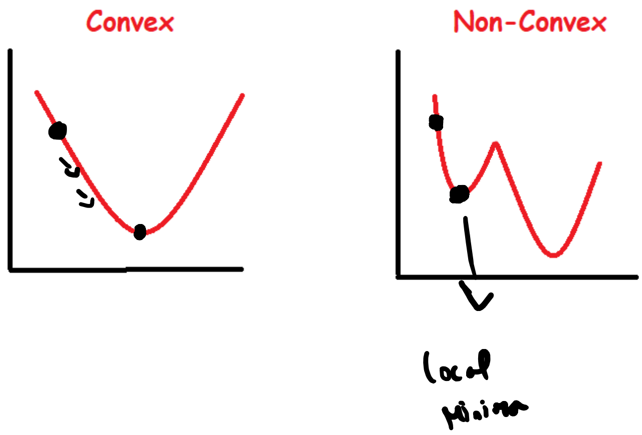


Deep Learning

→ local Minima ≠ global Minima



nonconvex
_loss_and...



Nothing

{[Stochastic G.D]
[Newton / Quasi Newton Methods]

Interview

[Which algorithm ensures a convex loss function? ML]

[Linear Regression]

Linear Regression

Should we repeat the process if loss function is same for a model

05 November 2025 15:04

$$\left[loss = -\sum \frac{(w^t x + w_0)\, y_i}{\|w\|} \right]$$

Do I need to apply G.D everytime?

$\left[10 \text{ different classification problems} \right]$

$$\sum \frac{\left(w^t \underline{x} + w_0\right) \underline{y_i}}{\|w\|}$$

$\left[\text{1000 data and 10 columns} \right]$

loss significantly

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}$$

x_6