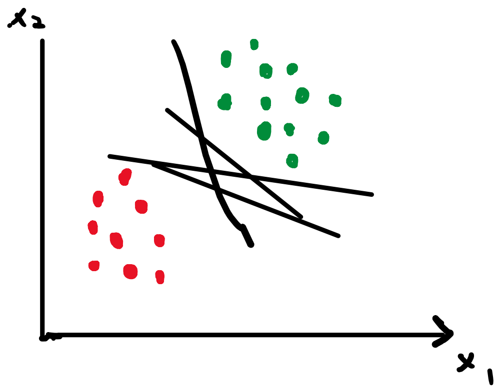
Need for loss function

27 August 2025 19:18



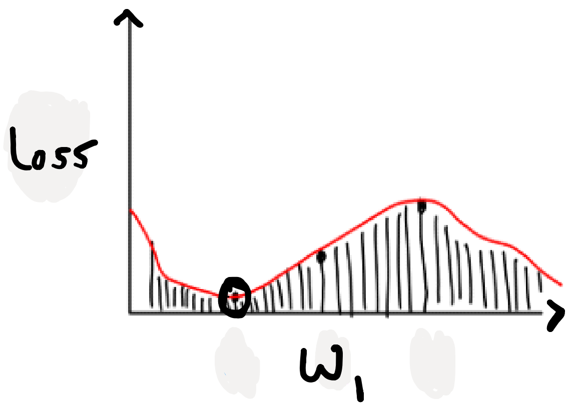
How Many Possible Lines can seperate this data?

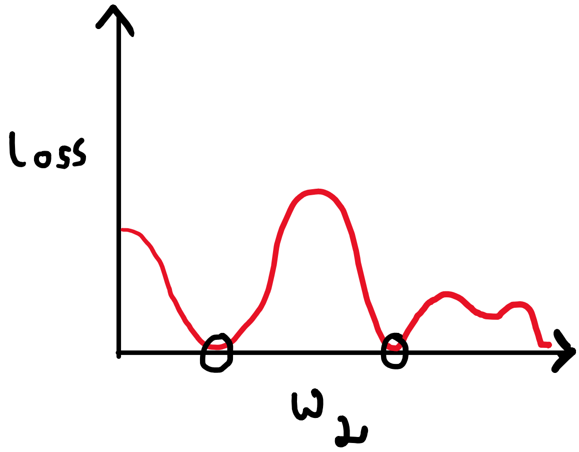
Infinity.

But we need one line. For that, we want to Minimize loss function

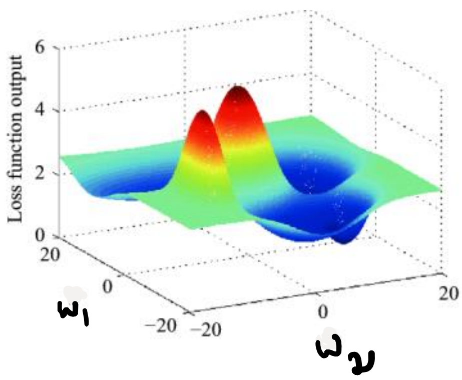
	w1	w2	w0	Accuracy (%)	loss
L1	0.5	-0.3	0.1	72	33
L2	-0.8	0.6	0.0	65	32
L3	1.2	0.9	-0.4	81	11
⋮	-1.0	-0.5	0.3	60	40
⋮	0.7	1.1	-0.2	78	15
⋮	-0.4	0.8	0.5	70	16
⋮	1.5	-1.2	-0.1	85	9
⋮	0.3	0.2	0.0	68	33
⋮	-1.3	1.0	0.2	63	37
⋮	0.9	-0.7	-0.5	77	14

If I plot one of the error vs one of the weights





Unique



Shape of loss function

Can we go through every possible sets of weights to find the best line?

Previous solution => Perceptron

[Lots of drawbacks]

Required solution :- Minimize Loss Function.

[Convex optimization]

limits

28 August 2025 12:47

Let us say a person is travelling from city A to City B. He travels such that distance remaining can be represented by $\boxed{d_{rem} = \frac{1}{2^t}}$

$1/2^t$

At $t = 0$

$A \longleftrightarrow B$

What is d_{rem}? 100% or 1

At $t = 1$

$1/2^t$

$A \longleftrightarrow B$

What is d_{rem}? $\frac{1}{2^1} = 50\%$

At $t = 2$

$A \longleftrightarrow B$

What is d_{rem}? $\frac{1}{2^2} = \frac{1}{4} = 25\%$

At $t = 5$

$A \longleftrightarrow B$

$t = 10$

What is d_{rem}? $\frac{1}{[illegible]} = \frac{1}{[illegible]} = 3.2\%$ (1 2

What is d_{rem} ? $\frac{1}{2^5} = \frac{1}{32} = 3.2\%$ $\left\{\frac{1}{2^t}\right\}$

[Will d_{rem} ever reach exactly 0 ?]

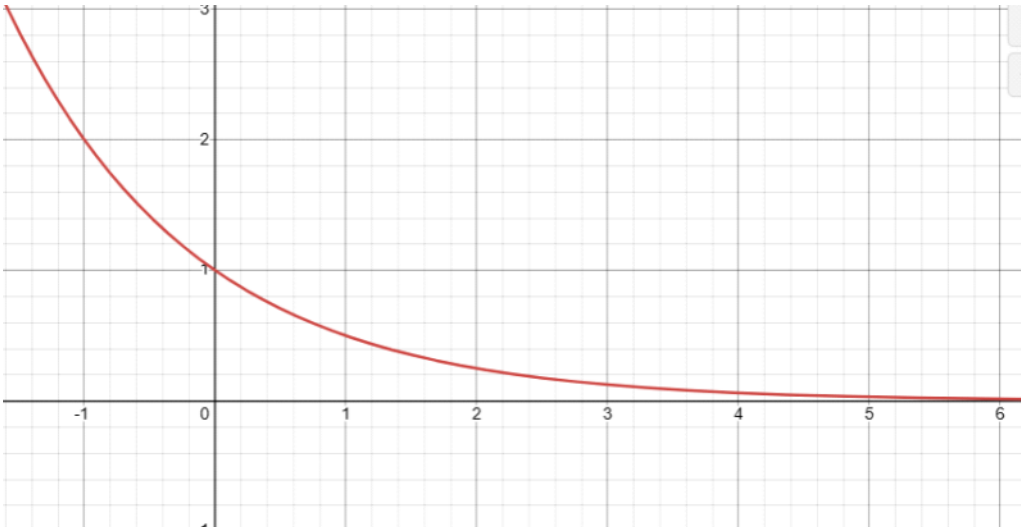
Close to 0 ? $\Rightarrow$ Yes !!

[However, will the driver eventually reach city B ?]

Yes !!

$1/2^t$

(limits) $\left[\lim_{t \to \infty} \left(\frac{1}{2^t}\right) = 0\right]$ $1/2^{1000}$



Take value of $\frac{1}{2^t}$ where t is close to 2, from [left hand side] $t = 1.999$

$$\frac{1}{2^{1.999}} = 0.25017 \approx 0.25$$

left hand limit $\Rightarrow \lim_{t \to 2^-} \left(\frac{1}{2^t} \right)$

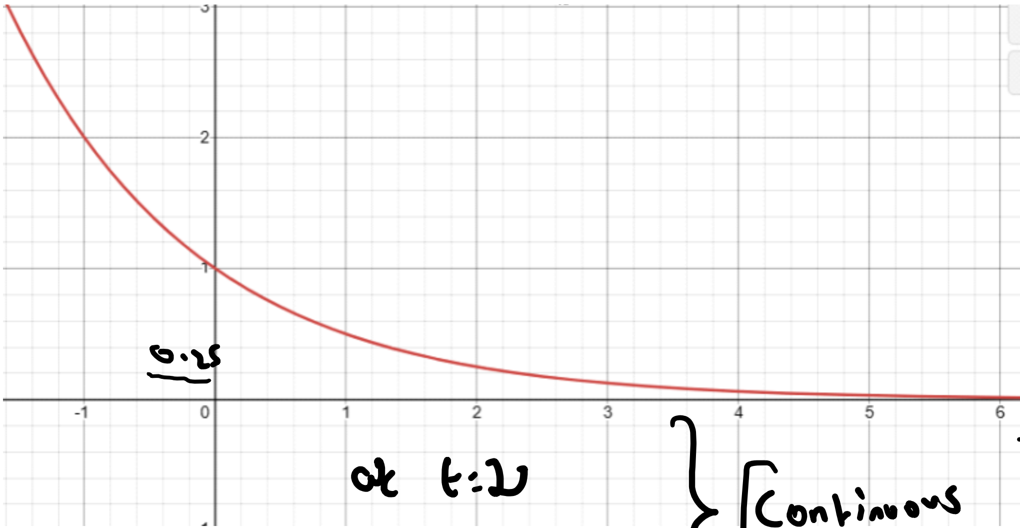
Take value of $\frac{1}{2^t}$ where t is close to 2, from right hand side

$$\frac{1}{2^{2.001}} = 0.2499 \approx 0.25$$

Right hand limit $= \lim_{t \to 2^+} \left(\frac{1}{2^t} \right)$

[Continuity]

[Lhl = Rhl]



[lhl = Rhl]

[function is Continuous at that point]

at t=2

fn is continuous

[Continuous fn at all points]

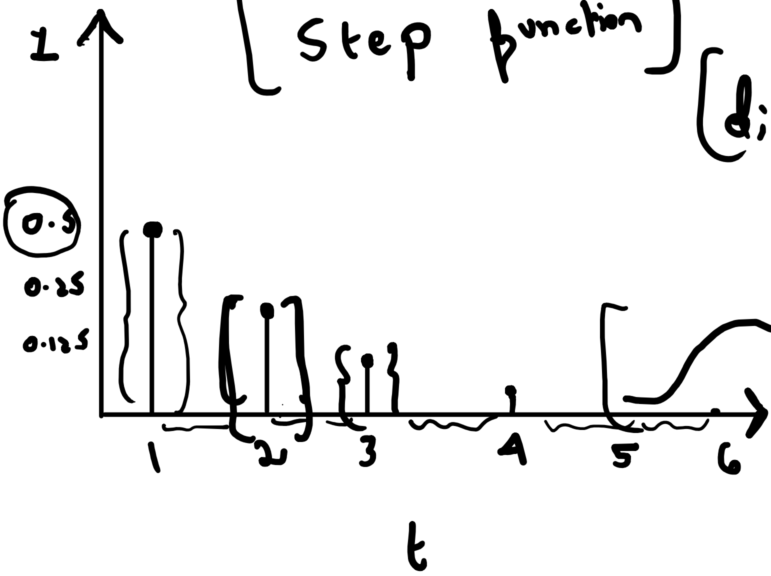
Lets say we dont have continuous info about the driver. We only get info about the driver when t is a whole number.

[Train Schedules]

Rail tracking

[Step function]

[discontinuous]



t = 2

[t = 1.999]

t = 2.001

[lhl = 0.5
Rhl = 0.25]

$$\lim_{t \to 2^-} f(t) = 0.5$$

lhl ≠ Rhl

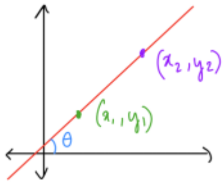
$$\lim_{2^+} f(t) = 0.25$$

function is discontinuous at $t = 2$

$$\begin{Bmatrix} \text{fn dis} & t < 5 \\ \text{fn cont} & t \geq 5 \end{Bmatrix}$$

Slope using limits?

30 August 2025 15:20

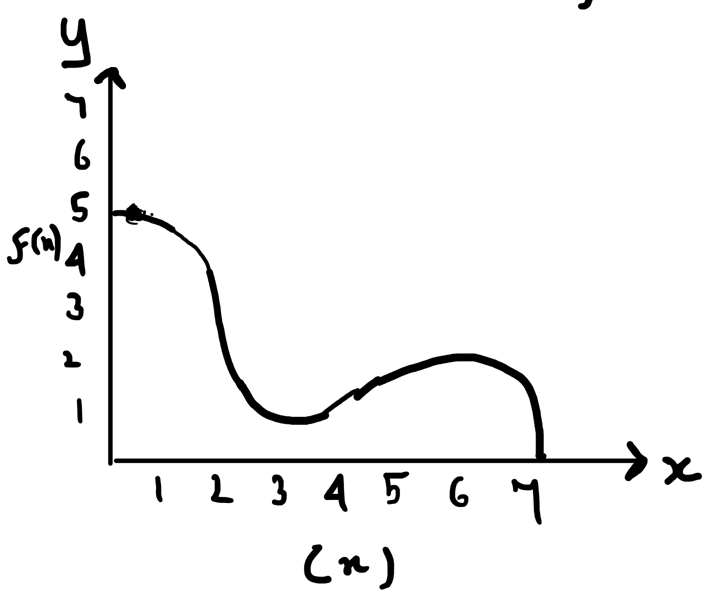


$m = \tan\theta$

$$\left\{ m = \frac{y_2 - y_1}{x_2 - x_1} \right\}$$

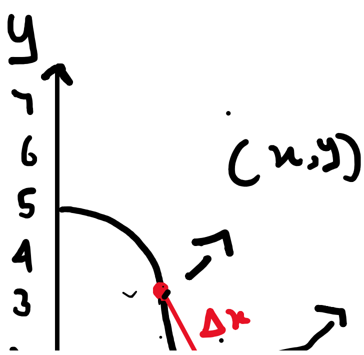
Does a curved line have a slope?

$y = f(x)$ y is a function of x.



How about slope between 2 points on a curve?

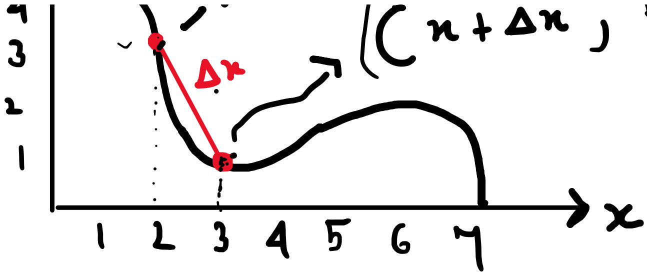
$$\frac{y_2 - y_1}{x_2 - x_1}$$



$$(x, y)\left\{\begin{array}{l} (\overset{x_1}{x}, \overset{y_1}{f(x)}) \\ (\overset{x_2}{x + \Delta x}, \overset{y_2}{f(x + \Delta x)}) \end{array}\right]$$

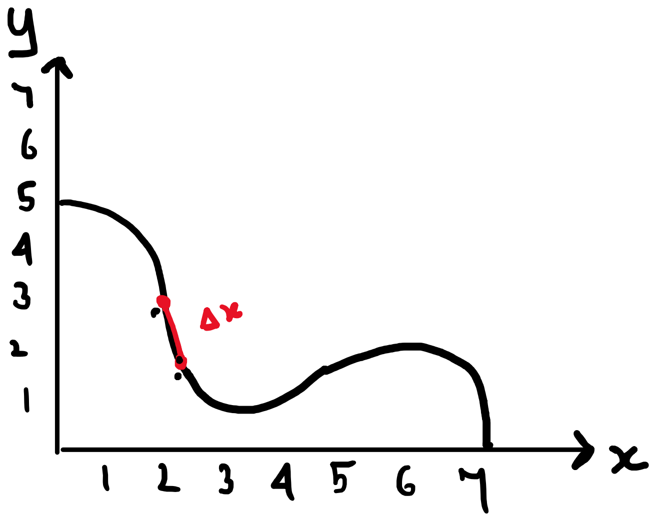
Slope = ?

$$\left[\frac{f(x + \Delta x) - f(x)}{x + \Delta x - \cancel{x}}\right]$$

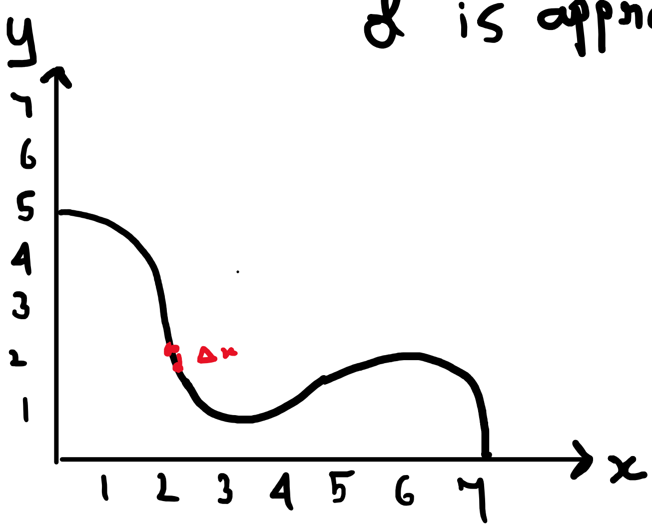


$(x + \Delta x, \ldots$

$\left(\dfrac{\ldots}{\cancel{x} + \Delta x - \cancel{x}} \right) = \dfrac{f(x+\Delta x) - f(x)}{\Delta x}$

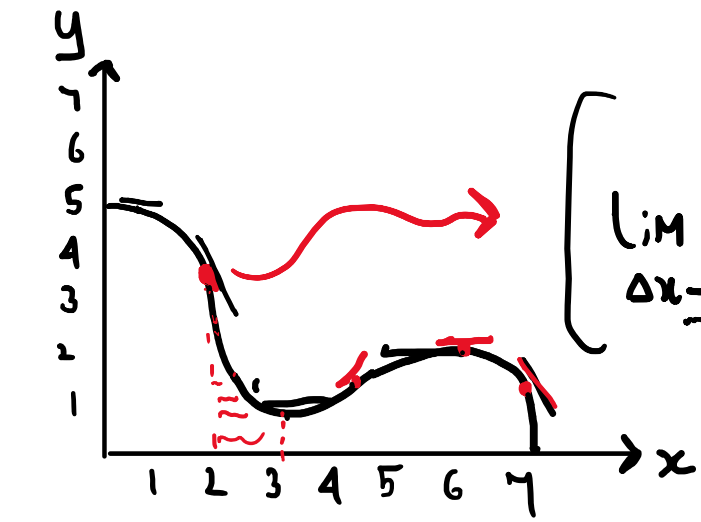


d is approaching 0



What value is distance approaching? 0

Derivative



$$\left[\lim_{\Delta x \to 0}\left(\frac{F(x+\Delta x) - F(x)}{\Delta x}\right)\right]$$

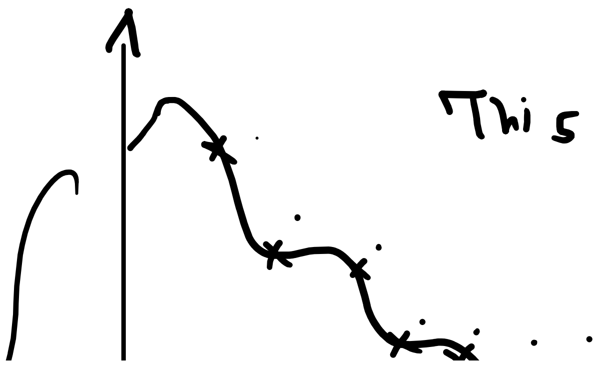
$\Delta x = 0$?

At a particular point, how is y changing with x

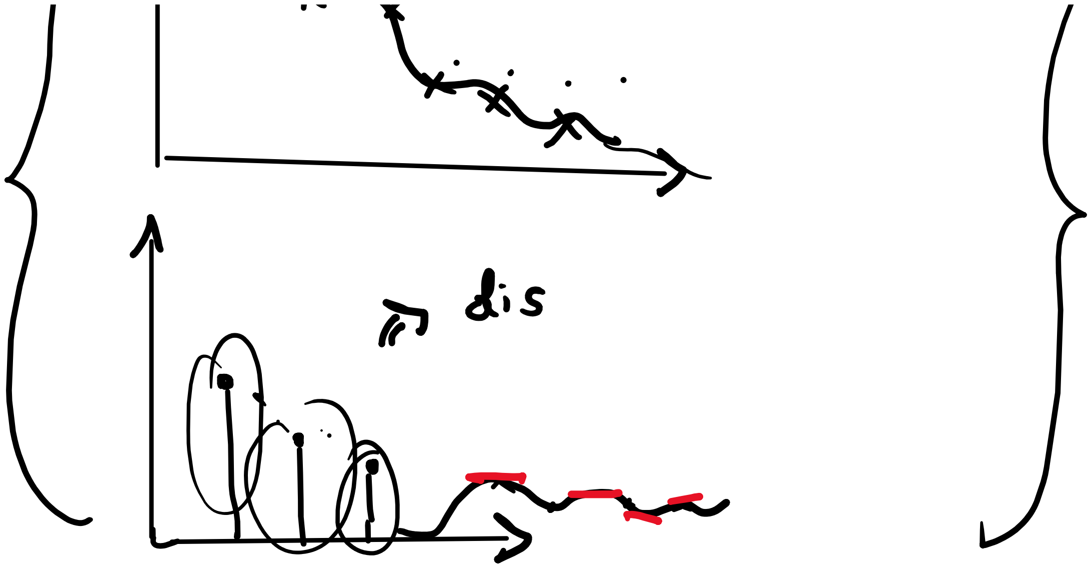
$$\text{Slope} = \left\{\frac{y_2 - y_1}{x_2 - x_1}\right\}$$

$\frac{dy}{dx}$ => rate of change

[A function can only be differentiable if it is continuous at a point.]



This is differentiable at all points



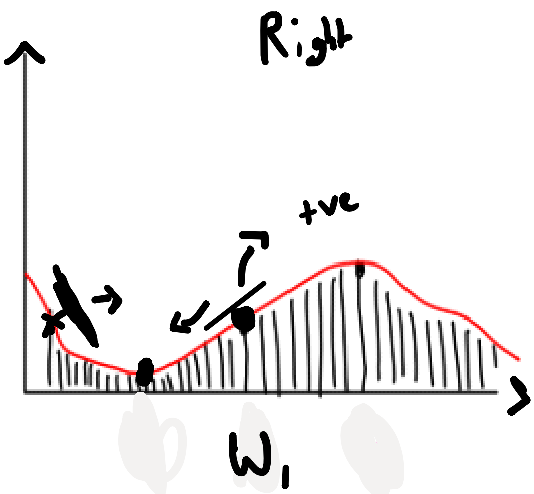
[Convex Optimization]

Minimum of loss fn

Link to Loss function

30 August 2025 16:15

diff weights ⇒ diff loss



Finding the derivative in order to reach Minima

Sneak peak ⇒ Convex optimization

Rules for Derivatives

02 September 2025 22:20

[Derivative of constant] $= 0$

$$\frac{d2}{dx} = 0$$

$$\frac{da}{dx} = 0$$

$$\frac{d\,x^n}{dx} \Rightarrow n \times x^{n-1}$$

$$x^3 \Rightarrow 3x^{3-1} = 3x^2$$

$$x^5 \Rightarrow 5x^{5-1} = 5x^4$$

$$6x^3$$

$$6 \times 3 \times x^2 = 18x^2$$

$$x_1 + x_2 \Rightarrow \frac{dx_1}{dx} + \frac{dx_2}{d}$$

$5x^3 + x^2$ Addition rule

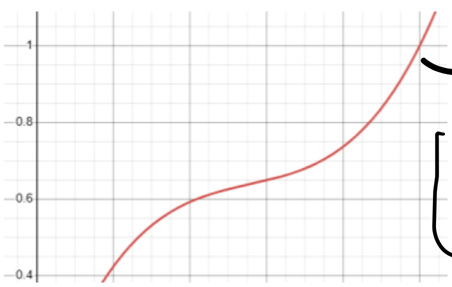
$$5 \times 3 \times x^2 + 2x^1 = 15x^2 + 2x$$

Multiply: $x_1 \times x_2 \Rightarrow x_1' x_2 + x_2' x_1$

Lets solve a real world case:

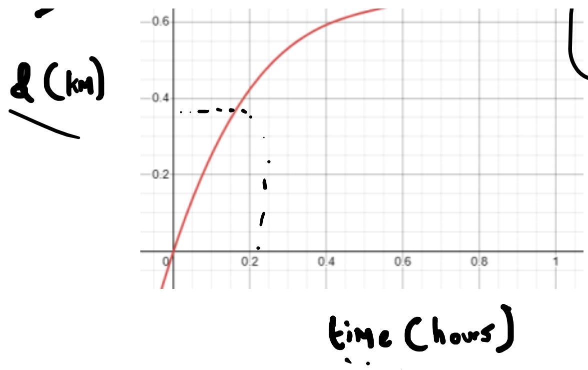
y

d (km)



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

[$y = 3t^3 - 5t^2 + 3t$]



distance $= (3t^3 - 5t^2 + 3t)$ $\quad t = 0.2$

[What was the dist at 0.2 hours ?] $\left[d = 3 \times 0.2^3 - 5 \times 0.2^2 + 3 \times 0.2 \right.$

What is the speed at $t = 0.2$

What was speed at 0.2 hours ?

$$s = \frac{d}{t} = \frac{\text{change in distance}}{\text{Change in time}} = \frac{\Delta d}{\Delta t} = \frac{\delta d}{\delta t}$$

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

speed: $9t^2 - 10t + 3$

Can I find speed at every point in time ?

Yes!

Chain Rule

30 August 2025 16:13

Multiple layers in a functions

$$(5x^2 + 2)^2 \qquad [\log(3x + 5x^2)]$$

Chain rule

Take inner function as another variable.

$$(5x^2 + 2)^2$$

$$5x^2 + 2 = u$$

$$\frac{du}{dx} = 10x$$

$$u^2 = 2u$$

$$2u \times (10x)$$

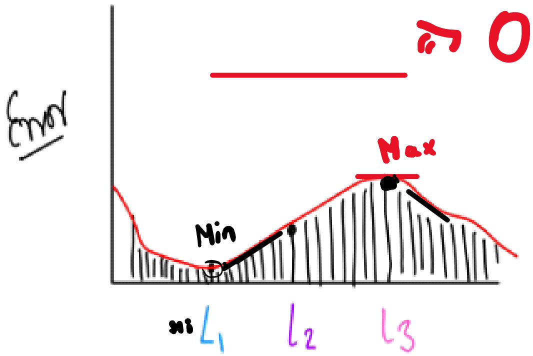
$$2(5x^2+2)\times 10x = 20x(5x^2+2)$$

Using Derivatives for Optimization

30 August 2025 16:18

[What is our goal in ML with respect to loss function?]

<u>Minimum</u> or Maximum



At both [Max, Min] Slope = 0

0 0

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

Max/Min

$8 - 3 \times 4 + 2$

$10 - 12$

$f(x) = 2$

$f(x) = -2$

$[3x^2 - 6x] = 0$

$3x(x-2) = 0$

$3x = 0$

$x - 2 = 0$

$[x = 0 \text{ or } 2]$

Critical points.

[Direction of slope] ⇒ Maxima/ Minima

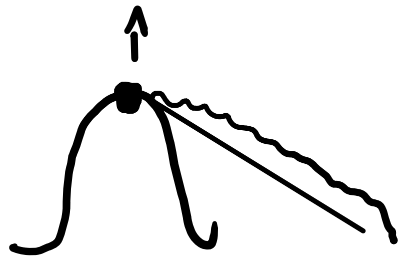
/ Minima

[derivative of the $\underset{\text{derivative}}{\text{slope}}$]

[derivative of the derivative] ⇒ double differentiation
2nd order derivative.

$$\{x=0\}$$

$$f''(x) = 6x - 6 \qquad x = 2$$



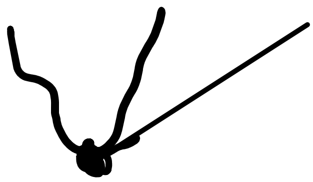
when $x = 0$ ⇒ Maxima.

$$f''(x) = -6$$

when $x = 2$

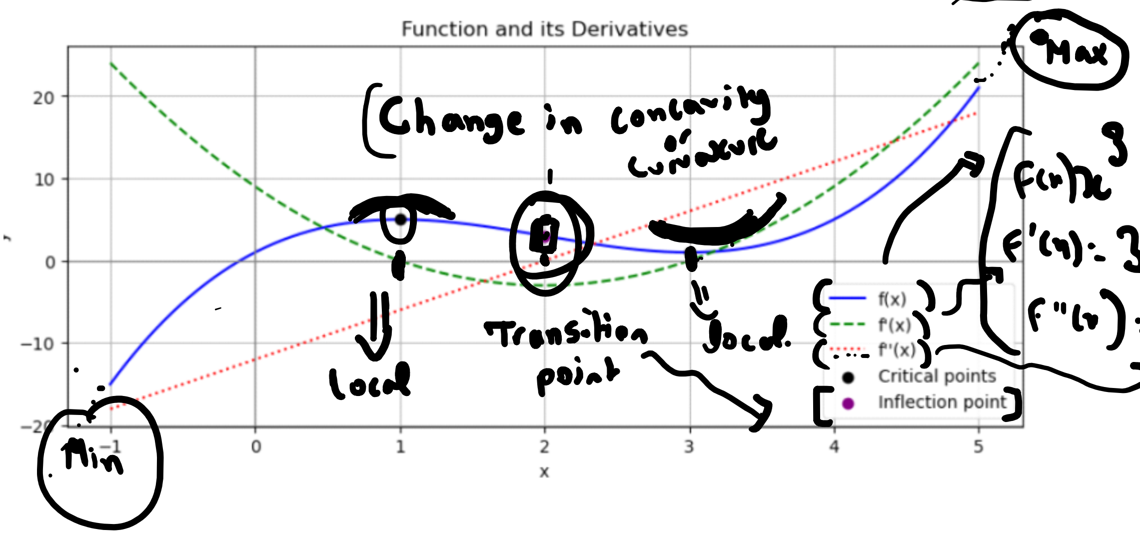
$f''(x) = 6$ ⇒ positive

[Minima]



ML ⇒ Second derivative as a [Hessian]

Convex optimization ⇒ [derivative + Hessian] to find optimal solutions.



$$f(x) = x^3 + 2x^2$$
$$f'(x) = 3x^2 + 4x$$
$$f''(x) = 6x + 4$$

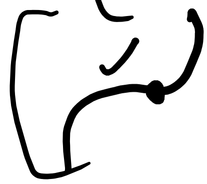
$$f''(x) = 0$$

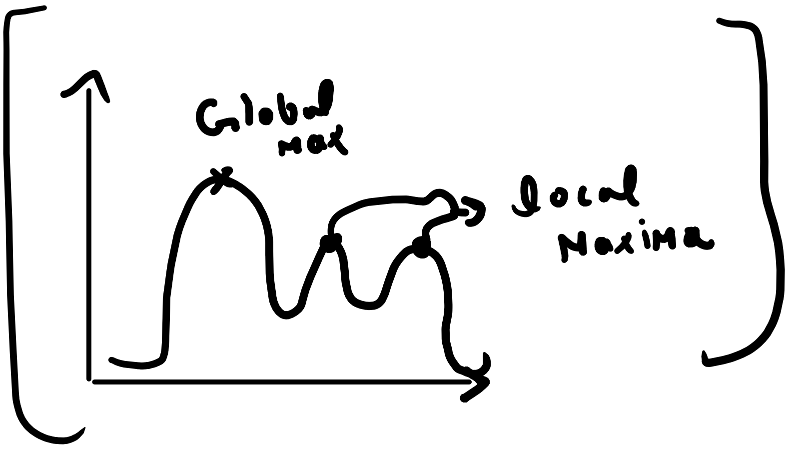
Hessian $= 0$

$$f'(x) = 0$$
$$f''(x) = 0$$

$$f'(x) \neq 0$$
$$f''(x) = 0$$

flex point





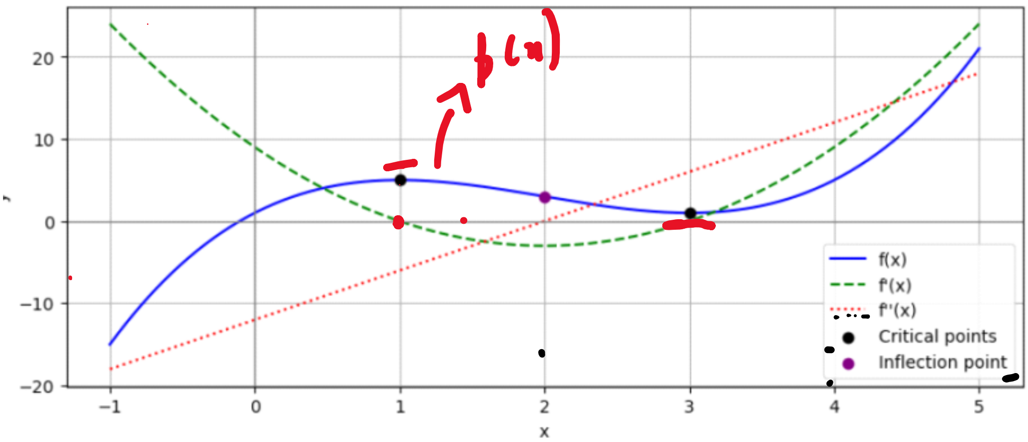
first derivative ⇒ local minima or local maxima

Global minima ?
or
Local Minima ?

Convex optimization

Next class

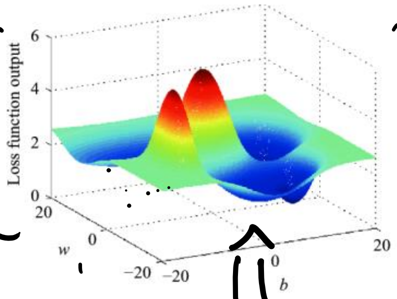
Local Min / Max



)

Why its not so simple

02 September 2025 22:38



First derivative

[Very very difficult]

$$\left[\log\left(w^2+w^3+e^{w^2}\right)\right]^3 \times e^{\log(x^2)}$$

Cannot be solved [analytically]?

$$\left[d \log\left(w^2+w^3+e^{w^2}\right)\right] \times e^{\log(x)} = 0$$

Never get a solution!!

Minima or Maxima.

$$\left[3x(x+2)=0\right]$$

System of Quad equation

$$x = 0$$
$$x = -2$$

$$\left[3x^3\left(\log(x^3)+e^{-x}\right]\right]^3 + \left(\log(e^{5x})\right)^3 = 0$$

$3x^{3}\,(\log(x^{3}) + e \;))\;(\;\; = 0$

[Convex optimization]