Introduction to Calculus

[Optimization]

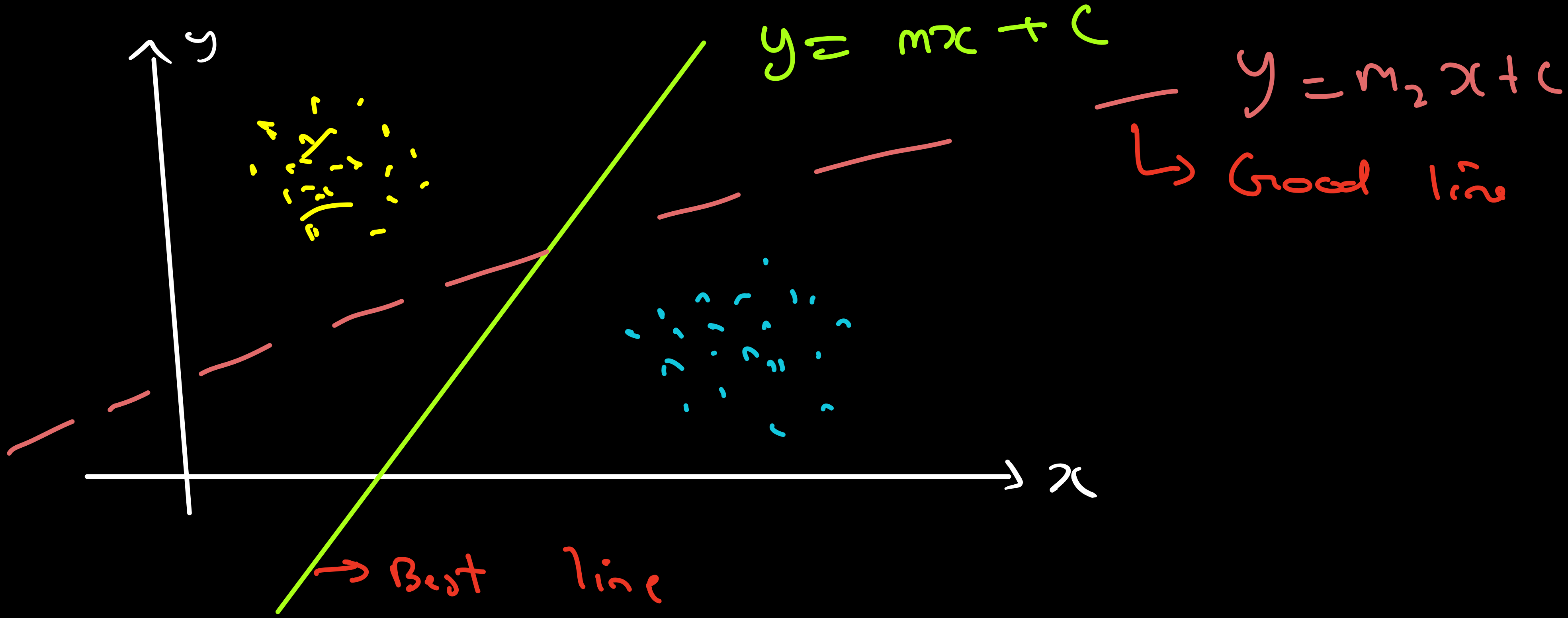
→ Motivation

→ Limits

→ Functions

→ Differenciation

Need for Calculus



We need to teach the computer to find the best line.

- First → How to determine / define the best line?

→ Distance of all point from the line should be maximum.

There are many points:

$$\begin{array}{c|cccc|} & f_1 & f_2 & f_3 & f_n \\ \hline \vec{x}_1 & x_{11} & x_{12} & x_{13} & x_{1n} \\ \vec{x}_2 & x_{21} & x_{22} & x_{23} & x_{2n} \\ \vec{x}_3 & \vdots & & & \vdots \\ \vdots & \vdots & & & \vdots \\ \vdots & x_{n1} & \cdots & \cdots & x_{nn} \\ \hline \end{array} \qquad \begin{array}{|c|} \vec{y} \\ \hline y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_n \\ \hline \end{array}$$

← Dataset

vectors

Features

Target

Mathematical definition for a binary classification

Given: $D = \{ (\vec{x}_i, y_i) ; \vec{x}_i \in \mathbb{R}^d , y_i \in \{-1, 1\} \}$

Find f such that,

$$f(\vec{x}_i) = y_i$$

OR

$$f(\vec{x}_i) = \hat{y}_i \sim y_i$$

model — We have to learn this

input

prediction / output

target / ground truth

Given dataset 'D', find 'f' such that $f(x_i)$ is close to y_i

We have decided that 'f' is going to be a function with the following formulation:

$$f(x_i) = \text{sign}\left(w^T x_i + w_0\right) = \hat{y}_i$$

We want $\hat{y}_i == y$ most of the times

prediction   ground truth

Now this all depends on $\underbrace{\vec{w}, w_0}$

"Machine Learning" refers to the act of automatically finding parameters. $\vec{w}, w_0$

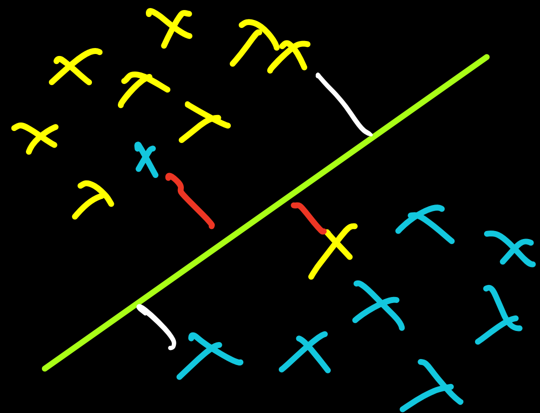
Best line

→ Distance of all points from the line should be maximum.

$$\text{distance} = \frac{w^T x_i + w_0}{\|w\|}$$

Q Can i say total distance should be max?

→ Yes



For miss classified points, however, i should negate it, since i dont want miss classifiн

+ve

−ve

Q: Can i interpret distance as confidence of prediction ?

a) No b) $D\uparrow = C\uparrow$ c) $D\uparrow = C\downarrow$

Mathematical Problem defⁿ

$$\boxed{\underbrace{\max_{\vec{w},\, w_0}}_{\text{optimisation parameters}} \underbrace{\sum_{i=1}^{n} y_i \cdot \frac{w^T x_i + w_0}{\|w\|}}_{\text{optimisation target / loss function}}}$$

optimisation parameters

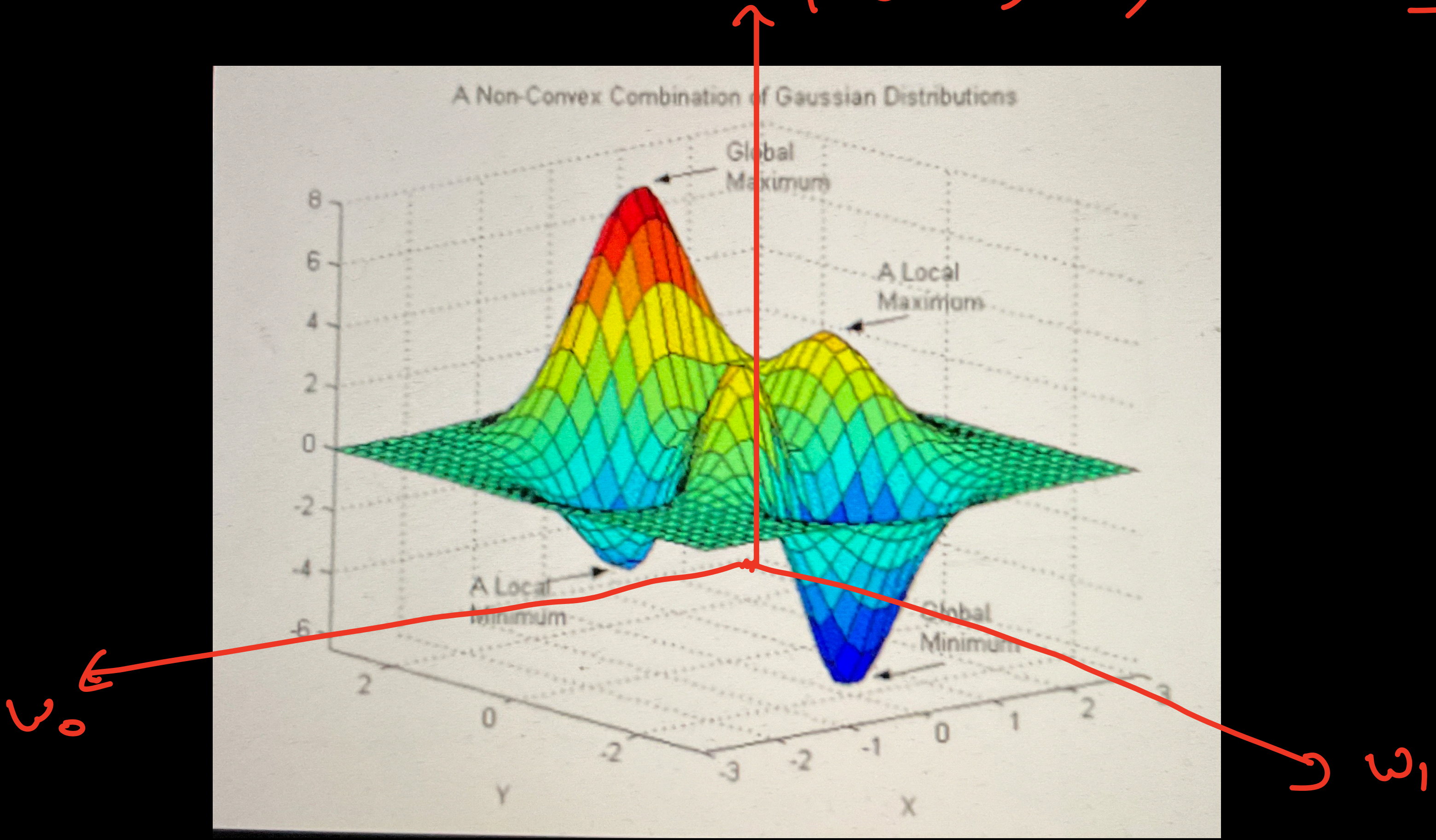
optimisation target / loss function

y_i is multiplied so that if

$y_i \neq \text{sign}(w^T x + w_0)$ we get −ve distance

Whichever values of $\vec{w}$ and w_0 maximise the opt target will give us the best line

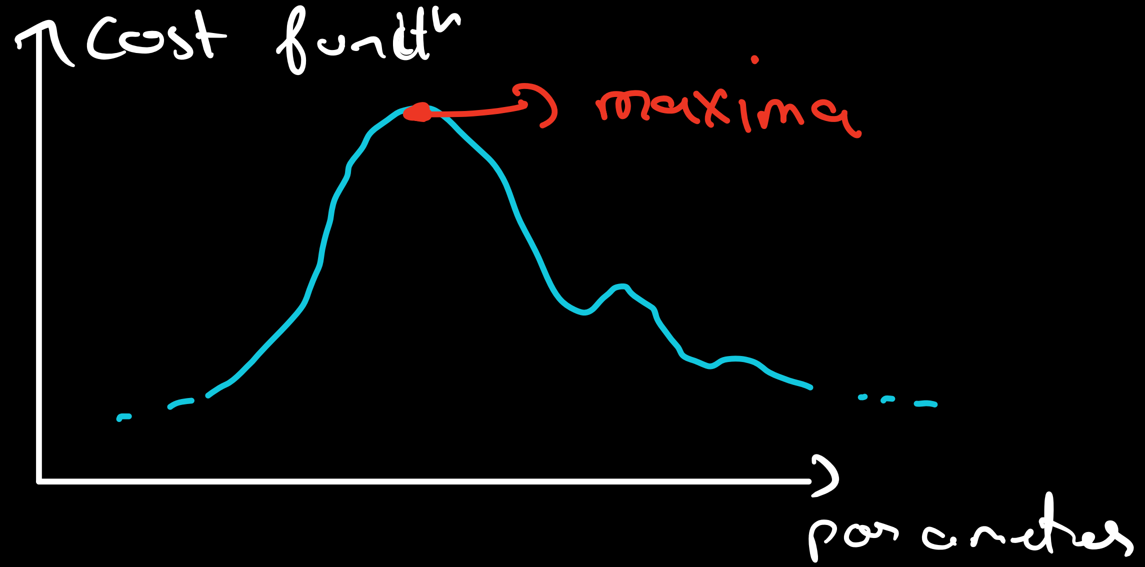
$F(w_1, w_0) = \sum y_i \frac{(w_1 x_1 + w_0)}{w_1}$



Note that this is a seperate space, here axes are $\vec{w}$, w_0, and

is seperate !!

In 2D



During "training" all the $\vec{x}_i$ and y_i are known.
$\vec{w}$ and w_0 are the parameters.

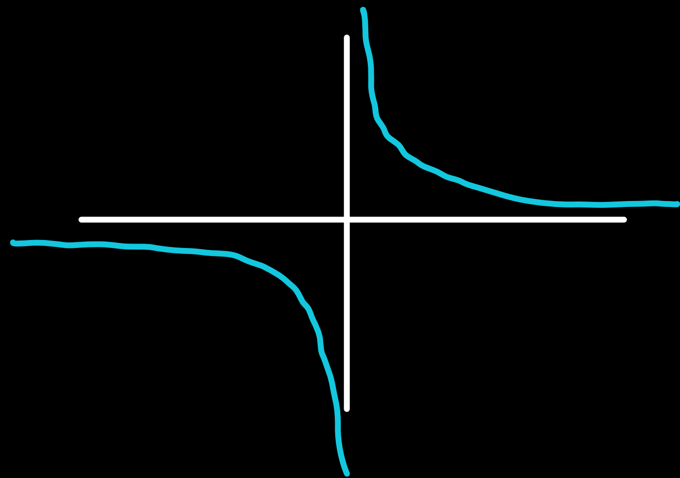
In order to understand, how to find the best parameters to optimise the value of loss function we need to know some basic calculus.

Limits

↳ The value of a function at extreme i/p

Eg:

$$f(x) = \frac{1}{x}$$ → desmos



$f(3) = 1/3$

$f(\infty) = \frac{1}{\infty} = 0$

$f(-\infty) = \frac{1}{-\infty} = 0$

$f(0) = ?$

$f(0^+) = \infty$ ⟶ $\lim_{x \to 0^+} f(x) = \infty$

$f(0^-) = -\infty$

$\lim_{x \to 0^-} f(x) = -\infty$

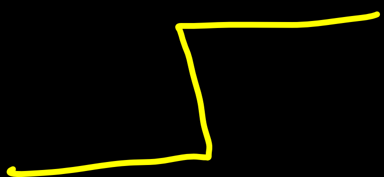
Q: What is limit of the following func^n

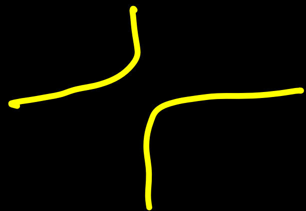
$$\lim_{x \to -\infty} e^x$$

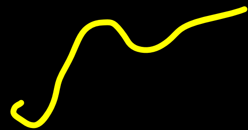
a) $-\infty$ b) ∞ c) 0 d) 1

Continuous Functions

Q: which of the following looks continuous?

a) 

b) 

c) 

A func^n is said to be continuous if LHS Limit = RHS Limit at each point

Eg: $f(x) = \frac{1}{x}$

RHS

$$\lim_{x \to 0^+} f(x) = \infty$$

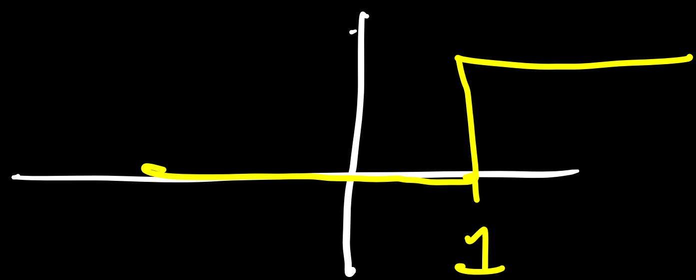
LHS

$$\lim_{x \to 0^-} f(x) = -\infty$$

LHS $\neq$ RHS

→ Not continuous

Eg: 2 Step function



$$\lim_{x \to 1^+} f(x) = \underline{\underline{1}}$$

$$\lim_{x \to 1^-} f(x) = 0$$ → Not continuous

Functions

A mathematical expression that converts an input to output is a function.

$x \rightarrow \underbrace{f(x)}_{\text{function of 'x'}} \rightarrow y$ | Univariate funcⁿ

$x, y \rightarrow \underbrace{f(x, y)}_{\text{funcⁿ of x, y}} \rightarrow z$ | Bivariate funcⁿ

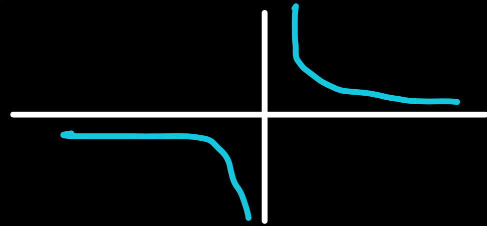
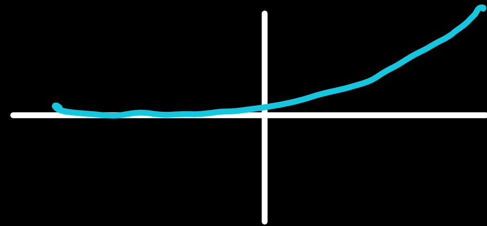
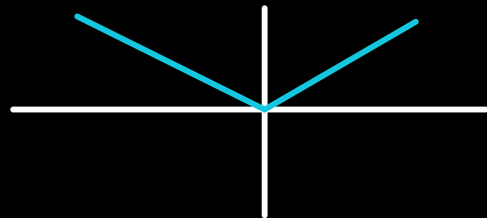
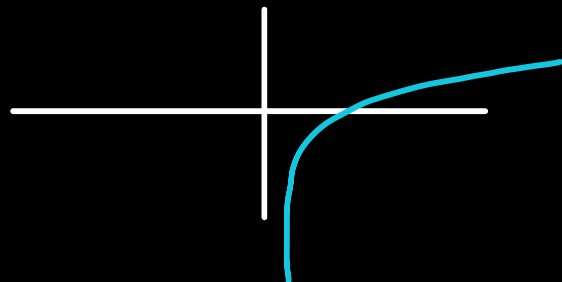
Domain

possible values as i/p

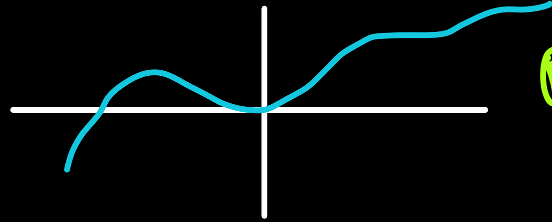
Range

possible values as o/p

Some imp functions for ML

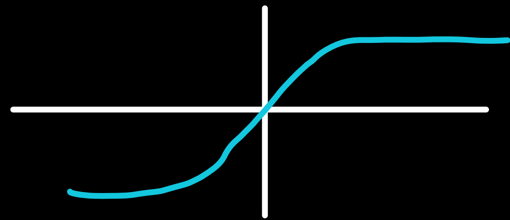
Domain	Range	Function	Plot	Name
$\mathbb{R}$	$\mathbb{R}$	$f(x) = \frac{1}{x}$		Hyperbola
$\mathbb{R}$	$\mathbb{R}^+$	$f(x) = e^x$		Exponent
$\mathbb{R}$	$\mathbb{R}^+$	$f(x) = \lvert x \rvert$		modulus
$\mathbb{R}^+$	$\mathbb{R}$	$f(x) = \log(x)$		logarithm

$\mathbb{R}$ $\quad$ $\mathbb{R}$ $\quad$ $f(x) = \sum_{i=1}^{n} a_i x^i$



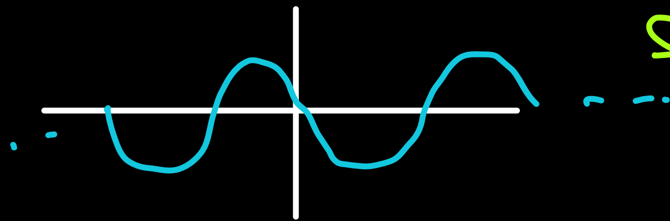
polynomial

$\mathbb{R}$ $\quad$ $(-1, 1)$ $\quad$ $f(x) = \dfrac{1}{1+e^{-x}}$



sigmoid

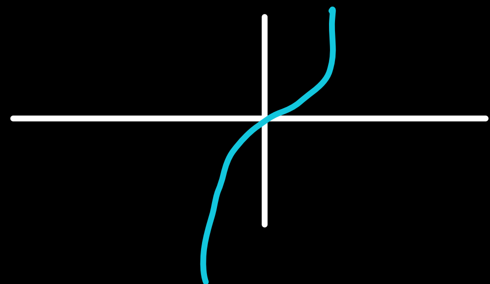
$\mathbb{R}$ $\quad$ $[-1, 1]$ $\quad$ $f(x) = \sin(x)$



sine

$\mathbb{R} - \underline{(2n+1)\frac{\pi}{2}}$ $\quad$ $\mathbb{R}$ $\quad$ $f(x) = \tan(x)$

odd multiples of $\frac{\pi}{2}$ not allowed



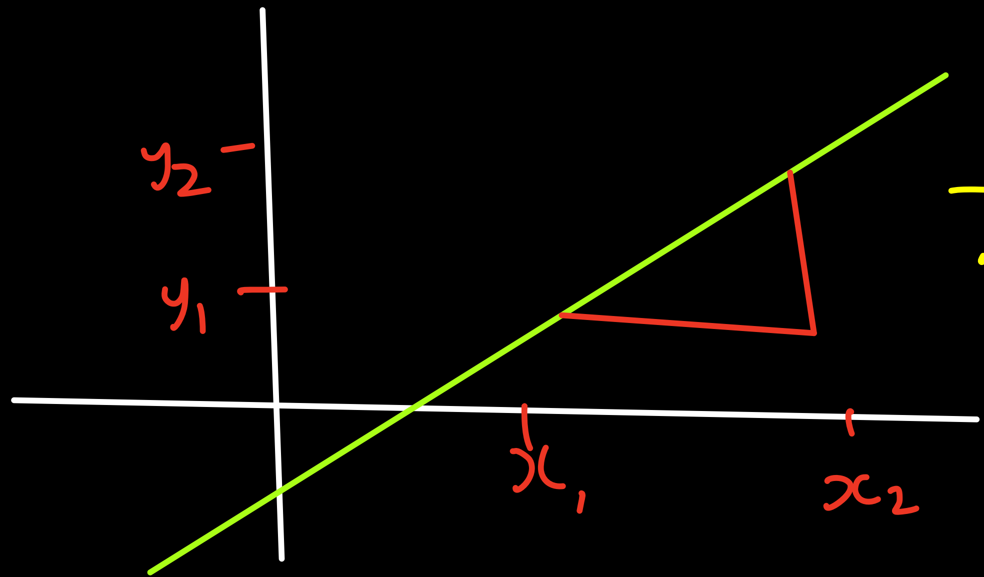
tangent

Differentiation

Let's take our fav func^n → $f(x) = mx + c$

$y = f(x) = mx + c$

How much will $f(x)$ change if I change 'x' ?



$$\frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}$$

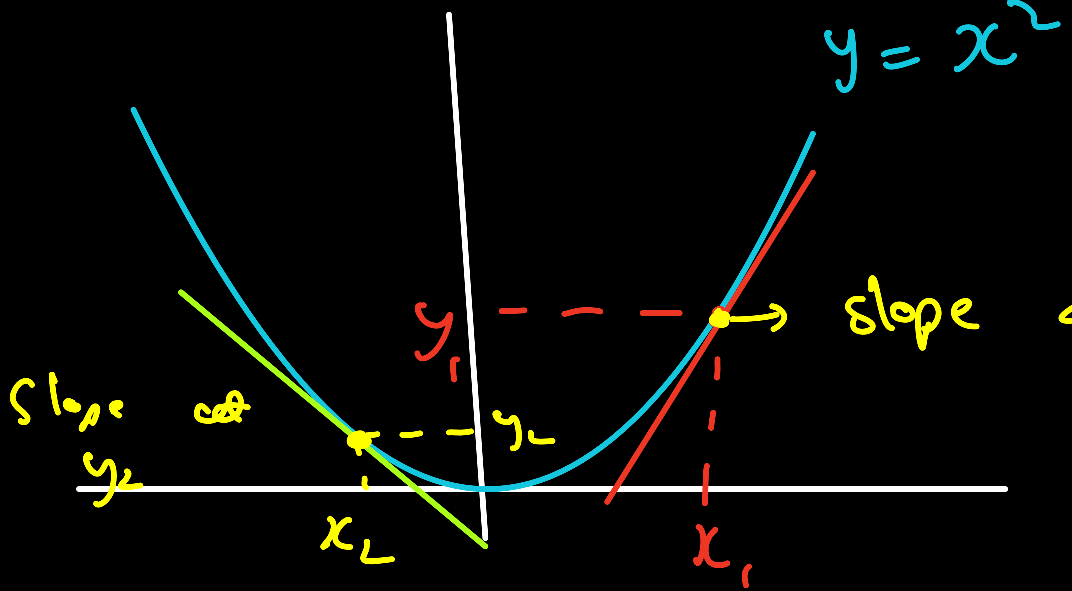
differenciation = rate of change = slope

$$\frac{d\,f(x)}{dx} = \frac{\Delta y}{\Delta x} = m$$

For $f(x) = mx + c$

$\frac{d\,f(x)}{dx}$ = differenciation of $f(x)$ with respect to x.

Let's take a slightly more complex func^n.



Here it is not correct to take $\frac{\Delta y}{\Delta x}$ because slope is continuously changing b/w $x_2 \to x_1$

Therefore we make $\to \Delta x \to \sim \underline{0}$

Very small change in $\underline{x}$

$$\therefore \frac{\Delta y}{\Delta x} = \frac{dy}{dx}$$

Now that we know, diff can be different at each point, we can define a new function $f'(x) = \frac{d}{dx} f(x)$

and some rules / formulas to find $f'(x)$

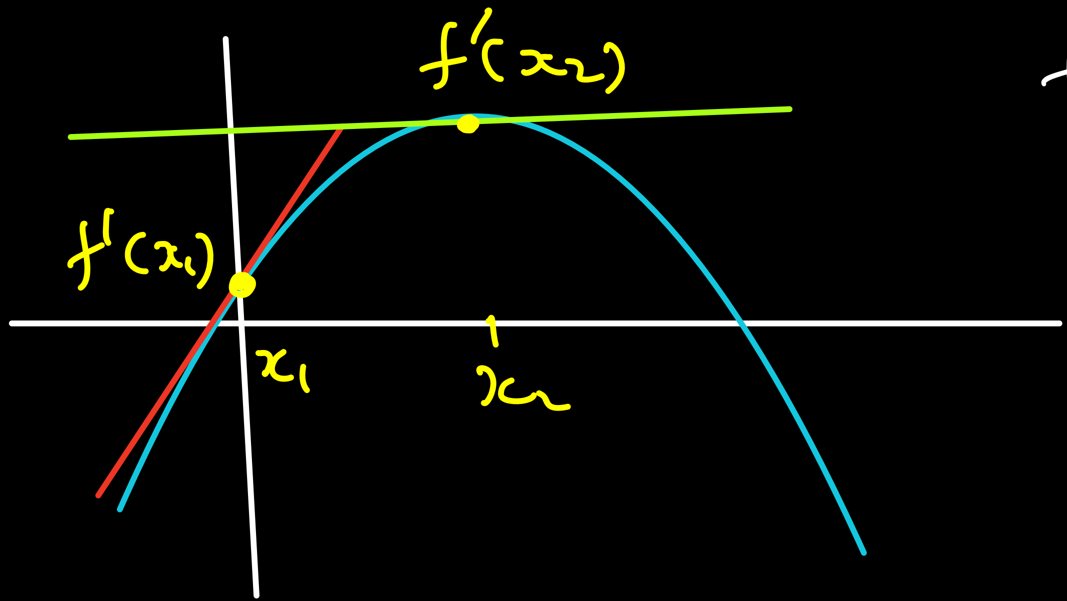
Some d/dx rules

$\frac{d}{dx} c = 0$	Constant Rule
$\frac{d}{dx} x^n = nx^{n-1}$	Power Rule
$\frac{d}{dx} \sin(x) = \cos(x)$	Trigonometric Rules
$\frac{d}{dx} \cos(x) = -\sin(x)$	
$\frac{d}{dx} b^x = b^x \ln(b)$	Exponential Rule
$\frac{d}{dx} \ln(x) = \frac{1}{x}$	Logarithmic Rule

Most important

Sum rule	$\frac{d}{dx}[f(x) + g(x)] = \frac{d}{dx} f(x) + \frac{d}{dx} g(x)$
Difference rule	$\frac{d}{dx}[f(x) - g(x)] = \frac{d}{dx} f(x) - \frac{d}{dx} g(x)$
Constant multiple rule	$\frac{d}{dx}[k \cdot f(x)] = k \cdot \frac{d}{dx} f(x)$
Constant rule	$\frac{d}{dx} k = 0$

Finding optima using Calculus



$$f(x) = a - 4(x-b)^2$$

$$f'(x) = 0 - 4 \times 2(x-b)$$

$$0 = -8x - 8b$$

$$\therefore \; x_{max} = \underline{\underline{b}}$$

$$\therefore \; f(x)_{max} = \underline{\underline{a}}$$

maxima $\rightarrow (b, a)$

$f'(x_1) > 0$

$f'(x_2) = \underline{\underline{0}} \rightarrow$ $\underbrace{\text{Maxima or Minima}}_{\text{optima / inflection point}}$ happens when $f'(x) = 0$

Using this math we will find max of our loss funcⁿ