It was being used even

Neural Networks (before Machine Learning)

1940 - Walter Pitts

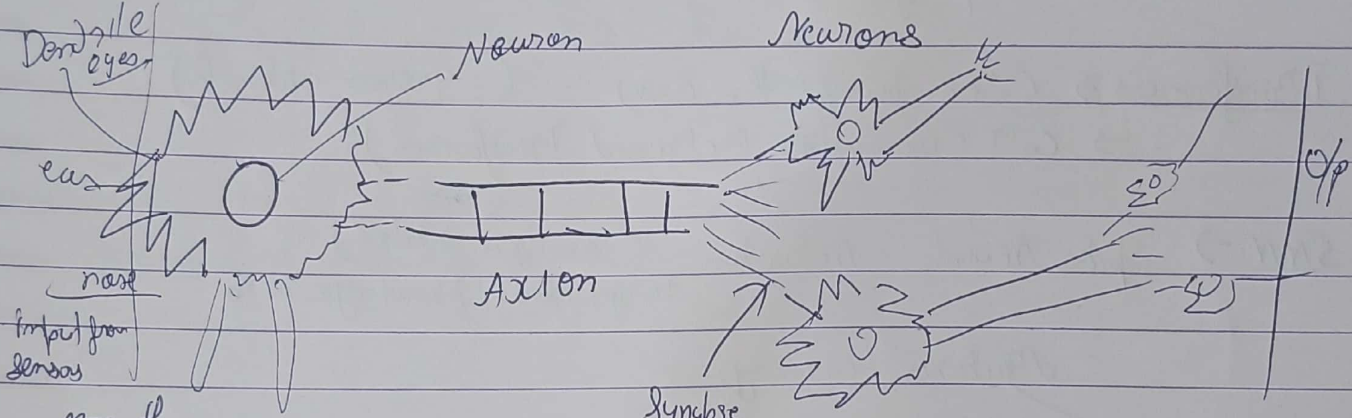
↳ What is the reason for Human Intelligence

→ Structure of Body → Robotics ⟨ ME, ECE, CSE

→ Good Brain → Machine Intelligence

└ CSE

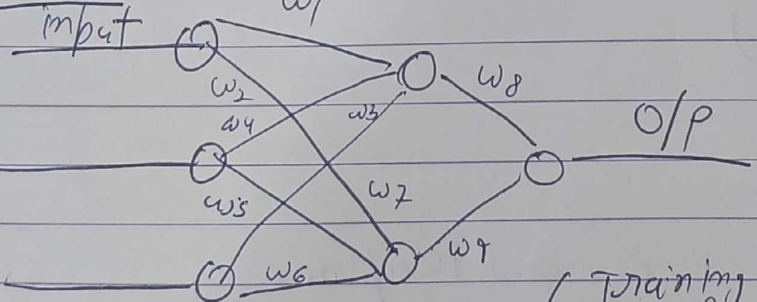
If we simulate the working of Human Brain → Neural Networks



More than 1000 Billion Neurons.

neurons work on training

Neural Network in CS



(Training is nothing but finding weights)

$$o/p = f(i/p, W)$$

FLA

Terms in NN

(Like Normal Neural Network)
(work on numbers)

FNN → Feed Forward Neural Netw.
CNN → Convolution Neural Network → Image processing
~~Im~~ RNN → Recurrent Neural Network
└ (based on previous data, predict next
autocorrect, grammarly, nlu

GAN → Generative Adversarial Networks → pure generate (content create)
(not deep fake (some input hota h)

Transformers → (More powerful than GAN) (queries chlani h)
↳ GPT (Generative Pretrained Transformer)

SNN → Spike Neural Network (mostly in robotics)
target class ki knowledge na ho

Machine Learning

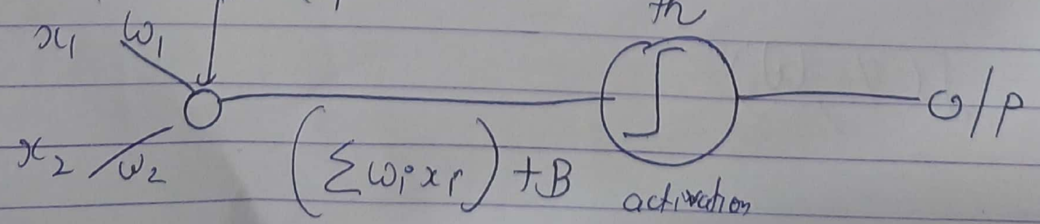
Supervised Learning
Classification
(Neural Network is part of Supervised Learning)
(training data diya target class bhi)

Unsupervised Learning
Clustering

Next milestone

1960 → Minsky → developed Perception

Perceptron



(Magnify Input)

activation function
↓
if input is greater than threshold then ____
else ____

Suppose AND Gate

x_1	x_2	O/P
0	0	0
0	1	0
1	0	0
1	1	1

$w_1 = 1$ $B = 0$ $th = 2$
$w_2 = 1$

If $k = x_1 + x_2 \geq 2$ then 1
else 0

Standard threshold 0 rkhte h
so $B = -1$ krdo

$x_1 + x_2 - 1 \geq 0$ the 1
else 0

OR Gate $B = 0$

XOR Gate (2 perception lene prega)

we work with SOP

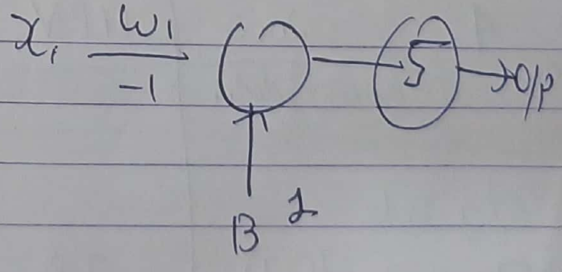
x_1	x_2	or POS
0	0	0
0	1	1
1	0	1
1	1	0

Interest with 1 o/p →

$T_1 + T_2$

$\bar{x}_1 x_2 + x_1 \bar{x}_2$

NOT



2005 → Geoffrey → Deep Learning

H/w Implement XOR Gate using perceptron

Basic formula

$$f_A \cdot (w_i x_i + \beta)$$

new wheels

wheel	weight	Target class		H.L
2	200kg	Bike/ motorcycle	→ 1	1
2	150kg	Motorcycle		1
4	800kg	LPV (Light Passenger vehicle)		1
4	50	Trolley		1
2	10	Bicycle		0
4	2000	MPV		1

(Training data)

Testing Data

2	100	??

(We cannot give this data directly into model)

→ We need to prepare our data

Data types in Neural Network

Qualitative → Binary, Nominal, Ordinal

Quantitative → Discrete, Continuous

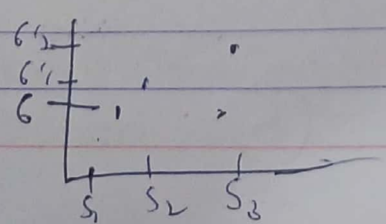
Binary
- 0/1
- Like in prev example if we add Head-Light

Nominal
- Only classes (no hierarchy no relationship)
- Class Phone company { };
  Apple, Android, Samsung, Nokia
  1, 2, 3, 4 (we attach any no.)
  no relationship b/w classes so no ordering of no.s as well

Ordinal
- There is relationship b/w classes
- Eg → Grades (A, A+, B, B+) → come from function & there is a order
- (there is a limit) / (no limit) → Discrete

Kitni classes bane pr discrete ho jayenge → Fuzzy Logic (Quantitative)

Point Chart ↳ Discrete Data



Line Chart → Continuous Data

→ we need to represent classes with no.s only
→ We cannot deal equally with 10 & 10,000 (need to normalize)

(2) Normalisation of data

i) Min Max Normalization

map it to

$$\text{wt from } [\overset{Min_o}{10} - \overset{Max_o}{2000}] \rightarrow (0-1) \text{ or } (A - B)$$

Formula

~~10 + 500 X (~~

$$\left\{ \frac{x - Min_{old}}{Max_o - Min_o} \right\} (B - A) + A$$

$$\left(\frac{500 - 10}{2000 - 10} \right) \times 1 + 0$$

$$\frac{49\cancel{0}}{199\cancel{0}} = 0.246$$

$$50 \rightarrow 60 \rightarrow 0 - 10$$

$$\boxed{x = 55}$$

$$\left(\frac{55 - 50}{60 - 50} \right) \times (10) + 0$$

$$= 5$$

ii) Decimal Scale Normalization

Suppose range is

Min_{old} —— Max_{old}

(ensure no negative no)

So most negative $\longrightarrow$ 0 a) (Add absolute value of most negative in everyone

b) or Change range acc to prev formula

Suppose highest no comes to be 999

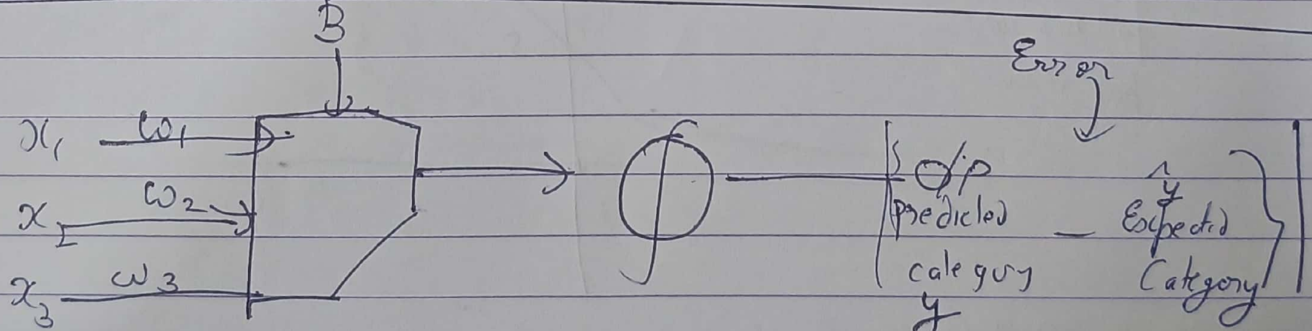
j min such that

Divide every no by (so that all no's align within 0-1)

$$\frac{x}{10^j}$$

☆☆

iii) Normalisation for genetic algorithms (work on binary no's)



$$\text{Error} = |y - \hat{y}|$$

BP

Neural Networking Algo & Back propagation will change wt to reduce errors.

$$\Delta\omega \propto |y - \hat{y}|$$

$$\omega = \omega + \Delta\omega$$

$$\beta = \beta + \Delta\beta$$

choti errors pe focus

i) Mean Absolute Error

$$(MAE) = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|$$

ii) Mean Squared Error

$$(MSE) = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

Big errors pe focus (outliers)

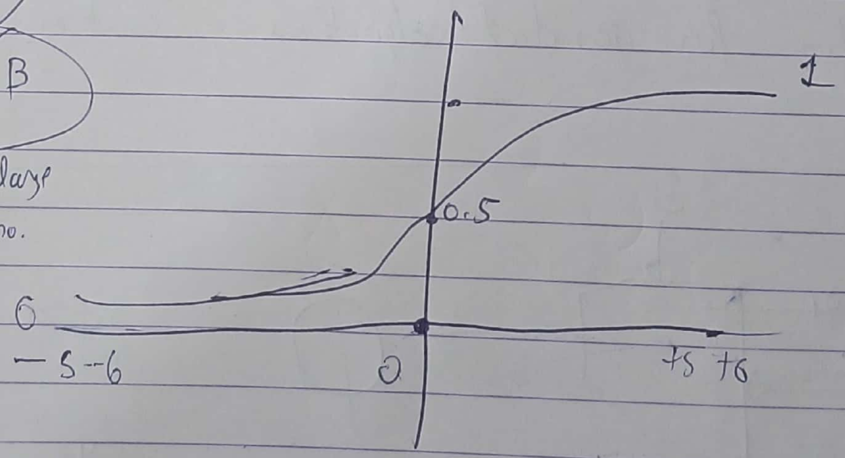
iii) RMSE (Root Mean Sq E) $= \sqrt{MSE}$

Most common Activation funct is Sigmoid / Logistic func^n

$$f(x) = \frac{1}{1+e^{-x}}$$

$\sum w_i x_i + B$

will be a large positive no.



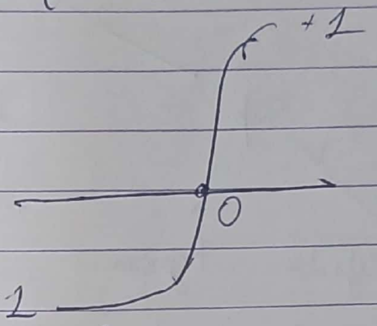
Gradients of Activation function

(Rate of change of wt will be corresponding to rate of change of errors)

$$MSE = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{1}{1+e^{-(w_i x_i + B)}} - \hat{y} \right)^2$$

tan h (tan hyperbolic) = ~~$e^x + e^{-x}$~~ $\dfrac{e^x - e^{-x}}{e^x + e^{-x}}$

~~Abx~~ better
(since aligned on 0)



rate of change of wt is prop to derivative of tan h

* Coefficient of Determinant (R)

$$R = 1 - \frac{\Sigma(y - \hat{y})^2}{\Sigma(y - \bar{y})^2}$$

Actual

y mean

Coefficient of Determination 1 means, the output is perfect
~~the~~ works in (when variance of x is in consideration

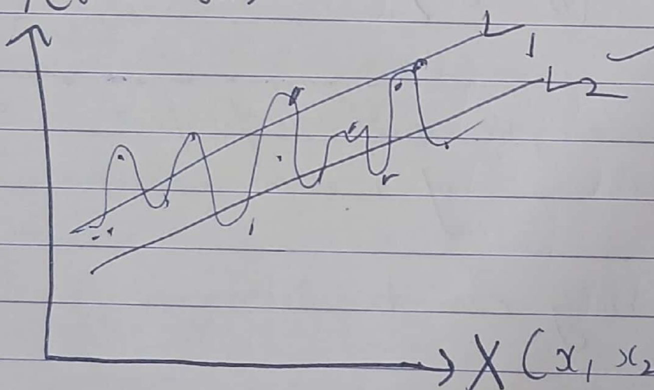
target class can be determined perfectly

(R=x) means
we say that x% of this variability can be determined

7th Feb/2025

ML
- Sup
  - Classification
    - Neural Net
    - Decision Tree
    - Random Forest
    - Naive Bayes
    - K-Nearest Neighbour (KNN)
  - Regression/Prediction
    - NN
    - Reg
      - Linear
      - Non Linear
    - SVM
- Unsup
  - Clustering
    - (K means)
    - DBScan
    - Hierar
- Ren
  - Q Algo

$Y(y_1 y_2 \dots y_n)$ $n <<< K$



$X(x_1 x_2 \dots x_K)$

we want to have a best fit Line (acc to these readings)

why this graph (ajeeb sa)

1. Actually Non Linear
2. Maybe Error

Principle of Least Square (Using PDE)

If you have suitable no of readings using PDE you can find a line such that error is minimum (pass through most of points)

Expansion

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!}$$

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(x)(x-a)^2}{2!}$$ → Taylor Series

$$\sum_{n=0}^{\infty} \frac{f^n(x) \cdot (x-a)^n}{n!}$$

Newton Changes this formula some how (used to find roots of poly)

Newton Raphson

→ Jab error ka function 0 hoga

~~Newton Webson~~

$$f(x+h) = f(x) + h f'(x) + \frac{h^2 f''(x)}{2!}$$

→ X If we discard higher order terms

put 0

$$0 \Rightarrow -\frac{f(x)}{f'(x)} = h$$

What is our line

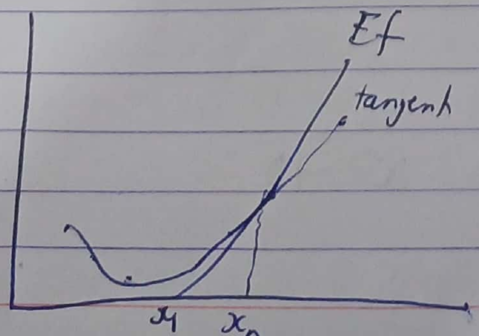
$y = a + bx$ → a & b are variable

So error function will be

$Ef = (y - (a + bx))$

Actual

Predicted



tangent at x_0 at which pt meets x axis is x_1 (Root ke bhut close proximity me h)

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} \leftarrow h \rightarrow$$

$0 = f(r)$ is f(r)th Root

$$x_1 = x_0 \cancel{t_1} - \frac{f(x_0)}{f'(x_0)}$$

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}$$

Eqⁿ of tangent

$$y - y_0 = \text{slope}\,(x - x_0)$$

$$y - y_0 = f'(x_0)(x - x_0)$$

where cutting x axis

$$y - f(x_0) = f'(x_0)(x - x_0)$$

this is close to 0 ~~(since near root)~~

~~$f(\text{nth}) \approx 0$~~

$$(x_1 - x_0) = -\frac{f(x_0)}{f'(x_0)}$$

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

When $f(x)$ will be 0

nya variable

pura be equal

hone lg jayega

wont deviate

Q) Evaluate $\sqrt{12}$ correct to 3 decimal places.

soln

Suppose $x = \sqrt{12}$

$x^2 = 12$

$x^2 - 12$

If we find the root of $f(x) = x^2 - 12$

$$f'(x) = 2x$$

when $x=3 \rightarrow f(x) = -ve$
$x=4 \rightarrow f(x) = +ve$

We will with close close approximate :- (3 or 4)

$x_0 = 3.5$

$$x_1 = 3.5 - \frac{f(x_0)}{f'(x_0)}$$

$$= 3.5 - \frac{(12.25 - 12)}{2 \times 3.5}$$

$$= 3.5 - \frac{0.25}{7}$$

$$= 3.5 - 0.0357$$

$$= 3.4643$$

x_1

$$x_2 = 3.4643 - \frac{(12.00137 - 12)}{2 \times 3.4643}$$

$$= 3.4643 - 0.00019$$

$$= 3.46411$$

$$x_3 = 3.46411 - \frac{(0.000058)}{2 \times}$$

$$= \underline{3.4641}$$

$$\frac{\sum_{i=1}^{N} \left(\left|y_i - (a+bx_i)\right|\right)}{N}$$

→ Sum of errors

→ mean error

mean Abs error

But there is problem of finding derivatives of absolute fn

So we use mean squared error

$$\frac{\sum_{i=1}^{N} \left(\left|y_i - (a+bx_i)\right|\right)^2}{N}$$

→ we are talking about variance

if error choti $2 \to 4$

$2 \to 10$
$\Delta = 8$

$(10 \to 100$

rate of change of error
$(\Delta = 96)$
↓
kafi bd gya

rate of change of error ⇒ derivatives

$$E(a,b) = \sum_{i=1}^{n} (y_i - (a+bx_i))^2$$

Err fn

(Ignore N for Now)

if we want to minimize Error

$$\frac{\partial E}{\partial a} = 0 \quad \text{and} \quad \frac{\delta E}{\delta b} = 0$$

to writ a

$$\left(2 \sum_i (y_i - (a+bx_i))(-1)\right) = 0$$

$$\Sigma y_i = na + b\Sigma x_i$$

— (1)

wrt b

$$2\sum y_i - (a+bx_i) \cdot (-x_i) = 0$$

$$\sum x_i y_i = a\sum x_i + b\sum x_i^2 \quad \text{— } 2$$

Comb 1 & 2

$$\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

this should be invertable (determinant non zero)

Determinant $\quad n\sum x_i^2 - \sum x_i \sum x_i$

$$n\sum x_i^2 - (n\bar{x})^2$$

$$n^2\left(\frac{\sum x_i^2}{n} - \bar{x}^2\right)$$

$\bar{x} = \frac{\sum x_i}{n}$

$\sum x_i = n\bar{x}$

→ Agr sari values same hogi toh zero ho jayega

$y = 2x + 3$

Suppose there change h

$\begin{bmatrix} a \\ b \end{bmatrix}$

x	y	$x_i y_i$
1	5	5
2	7	14
3	9	27
4	10.5	42
5	14	70
6	14.5	87
7	18	
	Σy_i	371

x^2

$\Sigma x^2 = 140$

$n = 7$

$\Sigma x_i = 28$

7 upon neeche obs main there fault

$78 = 7a + 28b$

$371 = 28a + 140b$

$371 = 140b$

$312 = 112b$

$= b = \frac{59}{28}$

$b = 2.107$

$a = 2.714$

$\Sigma y_i = 78$

$y = 2.107x + 2.714$

$x_i y_i$

31
28
18
96
77
58

we can similarly fit

$y = ax + bx^2 + c$

a & b & c be respect P.D.

& then solve

$$\begin{array}{r} 2 \\ 29 \\ \times 3 \\ \hline 87 \end{array}$$

What is formula of slope (b)

$$\sum y_i = na + b \sum x_i$$

$$a = \frac{\sum y_i - b \sum x_i}{n}$$

$$\sum x_i y_i = a \sum x_i + b \sum x_i^2$$

$$= \left(\frac{\sum y_i - b \sum x_i}{n}\right) \sum x_i + b \sum x_i^2$$

$$\sum x_i y_i = \frac{\sum y_i \sum x_i}{n} + b \left(\sum x_i^2 - \frac{(\sum x_i)^2}{n} \right)$$

$$b = \frac{\sum x_i y_i - \frac{\sum y_i \sum x_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}$$

$y = \sin x$

↳ Fourier Transformation

$y = ax + bx^2$

$$E_t = \sum_{i=1}^{n} \left(y_i - (ax_i + bx_i^2) \right)^2$$

$$\frac{\delta E}{\delta a} = 2 \left(\sum (y_i - ax + bx^2) \right) \cdot \left(-\sum x_i \right) =$$

4-5 Numerical (Linear Regression) → HW

Today's Topic: Curve Fitting using Principals of Least square

& Newton Raphson method of finding the roots

$$\frac{\delta E}{\delta a} = 0$$

$$\sum y_i x_i - a \sum x_i^2 - \cancel{c} b \sum x_i^3 = 0$$

$$\sum x_i y_i = a \sum x_i^2 + b \sum x_i^3 \quad (1)$$

$$\frac{\delta E}{\delta b} = 0$$

$$2 \sum y_i - (a x_i + b x_i^2 (-x_i^2)) = 0$$

$$\cancel{2} \sum y_i x_i^2 = a \sum x_i^3 + b \sum x_i^4 = 0 \quad (2)$$

x	y	$x_i y_i$	x_i^2	x_i^3	$y_i x_i^2$	x_i^4
1	5	5	1	1	5	1
2	15	30	4	8	60	16
3	25	75	9	27	225	81
4	47	188	16	64	~~3008~~ 752	256
5	68	340	25	125	1700	625
6	88	528	36	216	3168	1296
7	115	805	49	343	5635	2401
		1971	140	784	~~8377~~ 11545	4676

$$1971 = a\,140 + b(784)$$

$$\cancel{8377}\ 11545 = a(784) + b(4676)$$

Book I

Computer Based Numerical & Statistical Analysis

Linear Regression Analysis

Geometric & Algebric View

Vectors → Vector Spaces
→ Properties (Closure, Zero vector)

Linear Independence

Norm of a Vector

Rank of a Matrix

Basis of a Matrix

Imp

Dot Product
(Adjust weights for similarity)
Higher the value of dot product, higher similarity
→ Foundation for GPT
→ Recommendation System
Search Engines

PCA

ML Prerequisites

Partial Derivative

$$f_x(x,y) = \lim_{\Delta x \to 0} \frac{f(x+\Delta x, y) - f(x,y)}{\Delta x}$$

$$f_y(x,y) = \lim_{\Delta y \to 0} \frac{f(x, y+\Delta y) - f(x,y)}{\Delta y}$$

Total derivative

$$\frac{dz}{dt} = \frac{\delta f}{\delta x} \cdot \frac{dx}{dt} + \frac{\delta f}{\delta y} \cdot \frac{dy}{dt}$$

Power rule

$$\frac{d}{dx}\left(f(x,y)\right)^n = n\, f(x,y)^{n-1} \frac{d}{dx} f(x,y)$$

Product Rule

$$\frac{d}{dx}\left(f(x,y) \cdot g(x,y)\right) = f(x,y) \cdot \frac{d}{dx} g(x,y) + g(x,y) \cdot \frac{d}{dx}(f(x,y))$$

Quotient Rule

Chain Rule

Gradient of a function

↳ If we want to minimize the function (error) (reduce loss in ML) we move in the opp direct of the gradient

log loss

Imp