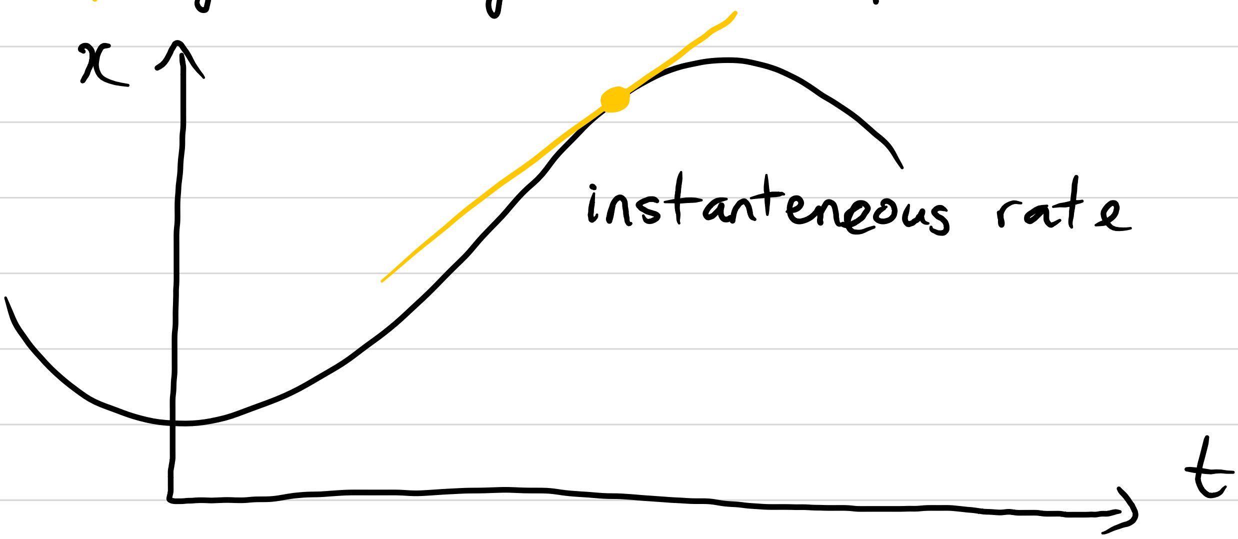
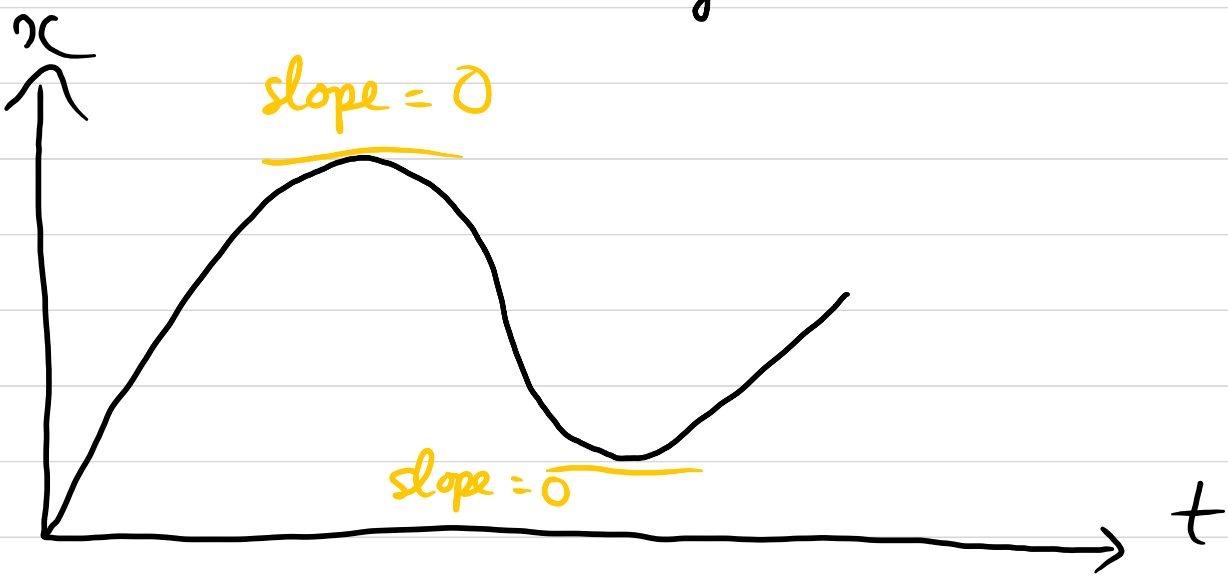
1. Derivatives

- Derivative of a function at a point is precisely the slope of the tangent at that particular point



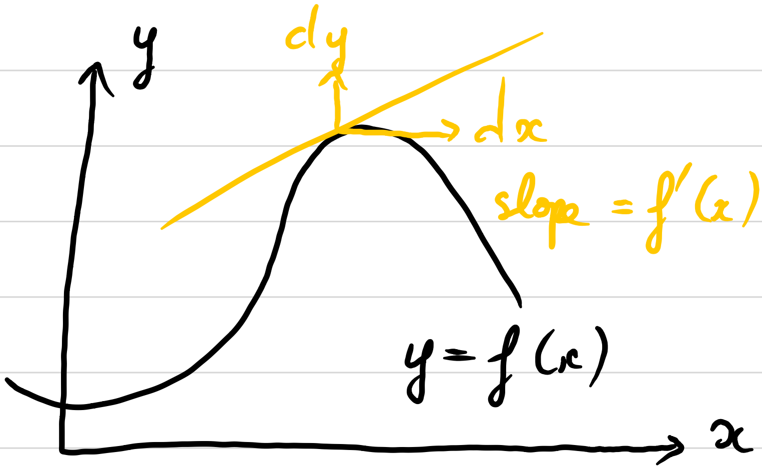
$$\text{slope} = \frac{\text{rise}}{\text{run}} = \frac{\text{change in } x}{\text{change in } t} = \frac{\Delta x}{\Delta t}$$



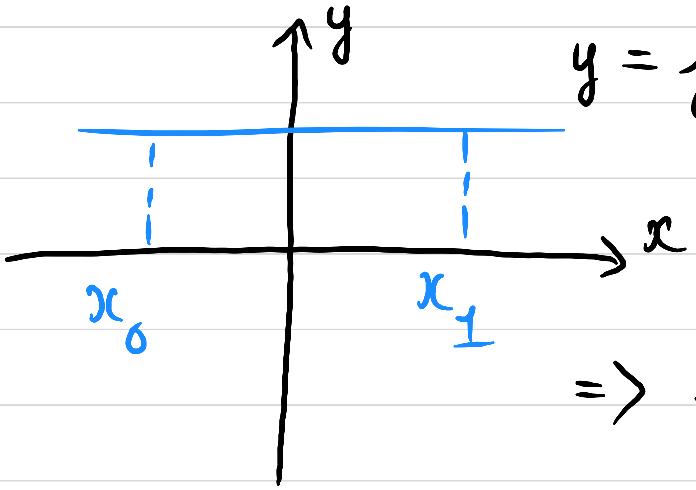
$y = f(x)$

$f'(x)$ Lagrange's notation

$\frac{dy}{dx} = \frac{d}{dx} f(x)$ Leibniz's notation



Derivative of a Constant

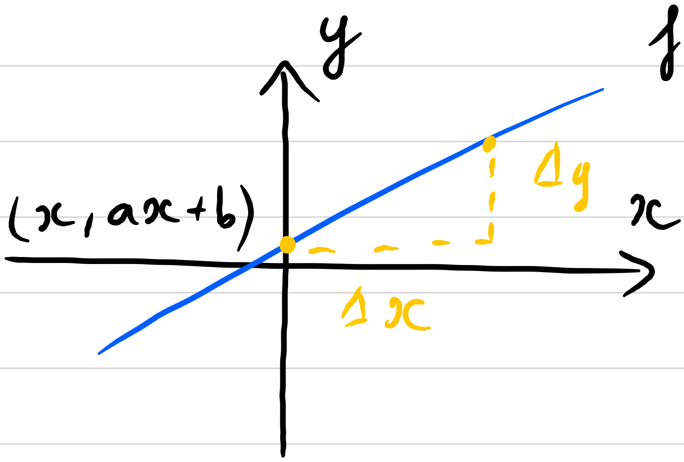


$$y = f(x) = c$$

$$\text{slope} = \frac{\Delta y}{\Delta x} = \frac{c - c}{x_1 - x_0} = 0$$

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

Derivative of a Line



$$f(x) = ax + b$$

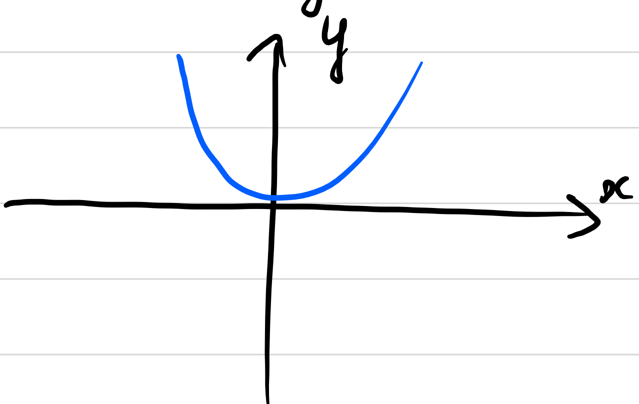
$$\frac{\Delta y}{\Delta x} = \frac{\text{rise}}{\text{run}} = a$$

$$= \frac{a(x+\Delta x)+b-(ax+b)}{x+\Delta x - x}$$

$$= \frac{a \cdot \Delta x}{\Delta x}$$

$$= a = f'(x)$$

Derivative of Quadratic Functions



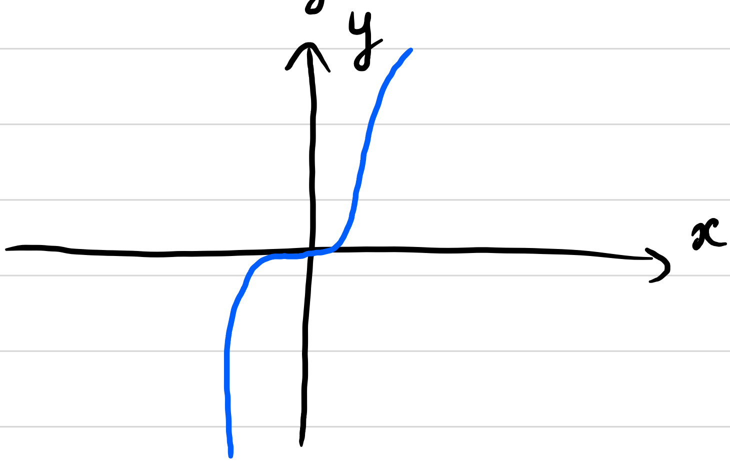
$$y = f(x) = x^2$$

$$\frac{\Delta f}{\Delta x} = \frac{(x+\Delta x)^2 - x^2}{\Delta x}$$

$$= 2x + \Delta x \quad (\Delta x \to 0)$$

$$= 2x$$

Derivative of Cubic Functions



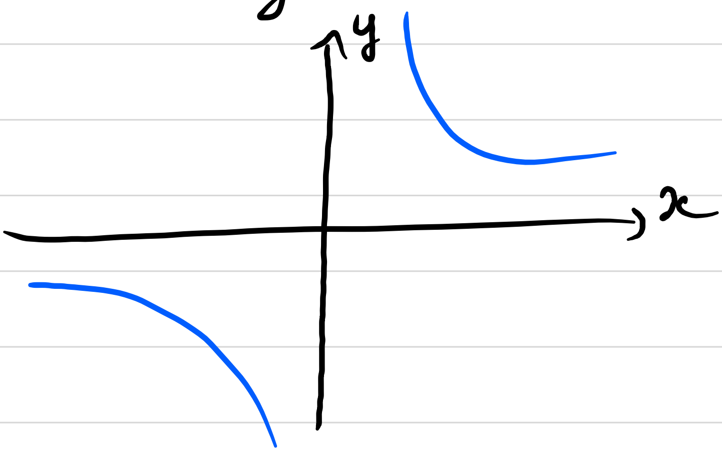
$$y = f(x) = x^3$$

$$\frac{\Delta f}{\Delta x} = \frac{(x+\Delta x)^3 - x^3}{\Delta x}$$

$$= 3x^2 + 3x\Delta x + \Delta x^2$$

$$= 3x^2 \quad (\Delta x \to 0)$$

Derivative of $\frac{1}{x}$



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

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

$$= \frac{-1}{x^2 + x\Delta x}$$

$$= \frac{-1}{x^2}$$

Inverse function

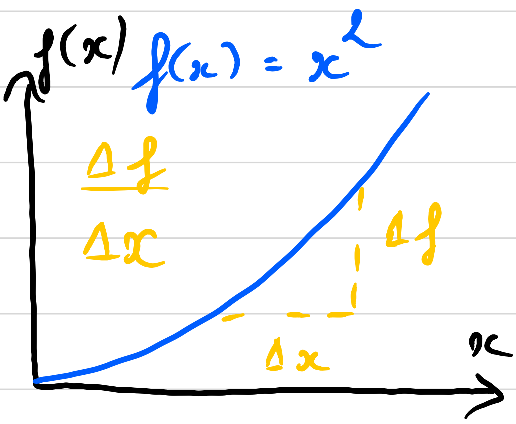
$$x \xrightarrow{f} x^2 \xrightarrow{g} x$$

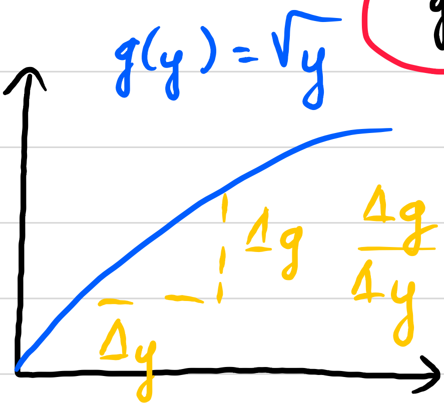
$f(x)$ $\qquad$ $g(x)$

x^2 $\qquad$ $\sqrt{x}$

$$g(x) = f^{-1}(x)$$

$$g(f(x)) = x$$





$$g'(y) = \frac{1}{f'(x)}$$

$$\frac{\Delta g}{\Delta y} = \frac{\Delta x}{\Delta f}$$

Derivative of Trigonometric functions

$$y = f(x) = \sin x$$
$$\Rightarrow f'(x) = \cos x$$

$$y = f(x) = \cos x$$
$$\Rightarrow f'(x) = -\sin x$$

Euler's number

$$e = 2.71828182 \qquad \left(1 + \frac{1}{\infty}\right)^{\infty} = e$$

$$f(x) = e^x \Rightarrow f'(x) = e^x$$

Derivative of $\log(x)$

$$e^{\log 3} = 3 \qquad e^{\log x} = x$$

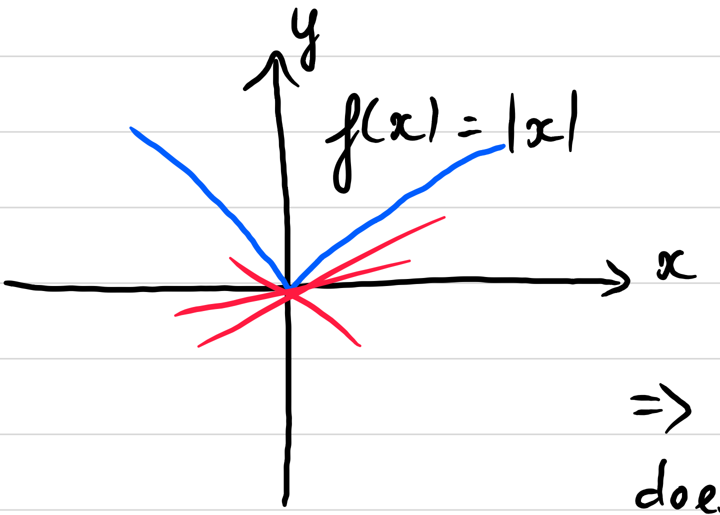
$$f(x) = e^x \Rightarrow f^{-1}(y) = \log y = g(y)$$

$\Rightarrow \log x$ is the inverse of e^x

$$f'(x) = e^x \Rightarrow g'(y) = \frac{1}{y}$$

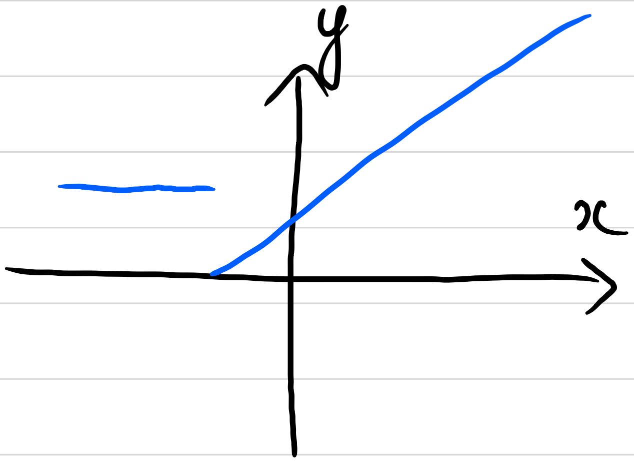
$$\Rightarrow \frac{d}{dy} \log y = \frac{1}{y}$$

Non differentiable functions



$$f(x) = \begin{cases} x, & \text{if } x \geq 0 \\ -x, & \text{if } x < 0 \end{cases}$$

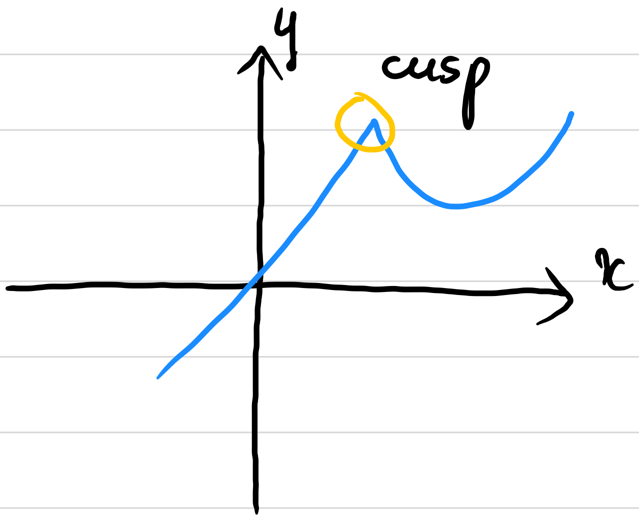
=> At $x = 0$, the derivative does not exist



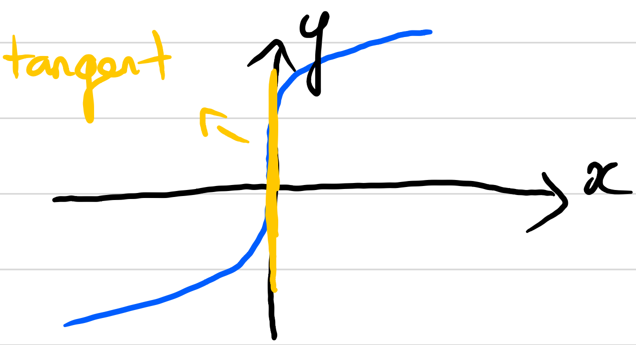
This is a piece-wise function

Jump Discontinuity

=> No continuous, not differentiable



=> Not differentiable



$$\text{slope} = \frac{\text{rise}}{\text{run}}$$

but run $= 0$ => Not differentiable

The product rule

$$f = gh \Rightarrow f' = g'h + gh'$$

The chain rule

$$\frac{d}{dt} f(g(h(t))) = \frac{df}{dg} \cdot \frac{dg}{dh} \cdot \frac{dh}{dt}$$

2. Optimization

- In ML, optimization is minimizing cost function

- One better way is using log loss

$$\frac{df}{dp} \rightarrow \frac{d}{dp} \log(f)$$

Why? $\Rightarrow$ Derivative of sum is easier than derivative of product

$$p^{16} \cdot (1-p)^{4}$$

$$\Rightarrow 16 \log p + 4 \log(1-p)$$

3. Gradients

Find the tangent plane

$$f(x, y) = x^2 + y^2$$

Fix $y = 4$ $\quad f(x, 4) = x^2 + 4^2$

$$\frac{d}{dx} f(x, 4) = 2x$$

Fix $x = 2$ $\quad f(2, y) = 2^2 + y^2$

$$\frac{d}{dy} f(2, y) = 2y$$

$\Rightarrow$ The tangent plane contains both tangent lines

Partial derivatives

$$f(x, y) = x^2 + y^2 \quad (\text{treat } x, y \text{ as constant})$$

$$\Rightarrow \quad \frac{df}{dx} = 2x \quad ; \quad \frac{df}{dy} = 2y$$

$$f(x, y) = 3x^2y^3$$

$$\Rightarrow \quad \frac{df}{dx} = 6xy^3 \quad ; \quad \frac{df}{dy} = 9x^2y^2$$

Gradient

$$\frac{df}{dx} = 2x \quad ; \quad \frac{df}{dy} = 2y$$

$$\nabla f = \begin{bmatrix} 2x \\ 2y \end{bmatrix}$$

Ex: $f(x,y) = x^2 + y^2$; calaulate ∇f at $(2, 3)$

$$\frac{df}{dx} = 2x \quad ; \quad \frac{df}{dy} = 2y$$

At $(2,3) \Rightarrow \nabla f = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$

Maxima/Minima (Optimization)

$$\frac{df}{dx} = 2x \quad ; \quad \frac{df}{dy} = 2y$$

Mininum when $\begin{cases} 2x = 0 \\ 2y = 0 \end{cases} \Rightarrow \begin{cases} x = 0 \\ y = 0 \end{cases}$

4. Gradient Descent

new point = old point − slope

$$x_1 = x_0 - f'(x_0)$$

→ $$x_1 = x_0 - \alpha f'(x_0)$$

learning rate

Function: $f(x)$ Goal: find minimum of $f(x)$

Step 1:

Define a learning rate α

Choose a starting point x_0

Step 2:

Update $x_K = x_{K-1} - \alpha f'(x_{K-1})$

Step 3:

Repeat step 2 until you are close enough to the true minimum x^* (converge) function: $f(x,y)$ goal: find min of $f(x,y)$

Step 1

Define α

Choose (x_0, y_0) starting point

Step 2:

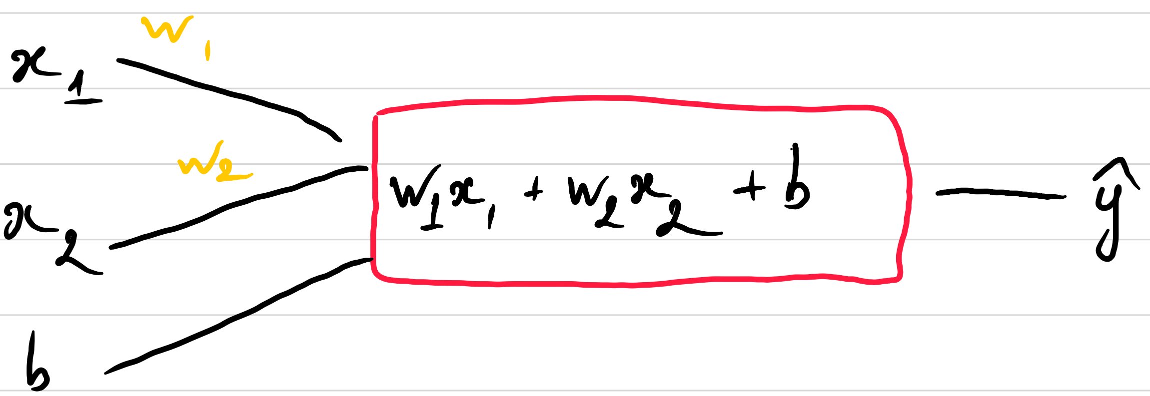
Update: $\begin{bmatrix} x_K \\ y_K \end{bmatrix} = \begin{bmatrix} x_{K-1} \\ y_{K-1} \end{bmatrix} - \alpha \nabla f(x_{K-1}, y_{K-1})$

Step 3:

Repeat step 2 until convergence

5. Optimization in neural networks

Regression with a perceptron



Main goal

- Find weights and bias that reduce prediction errors

Loss function

$$L(y,\hat{y}) = \frac{1}{2}(y-\hat{y})^2$$

To find optimal values for w_1, w_2, b :

Gradient descent

$$w_1 \rightarrow w_1 - \alpha \frac{dL}{dw_1}$$

$$w_2 \rightarrow w_2 - \alpha \frac{dL}{dw_2}$$

$$b \rightarrow b - \alpha \frac{dL}{db}$$

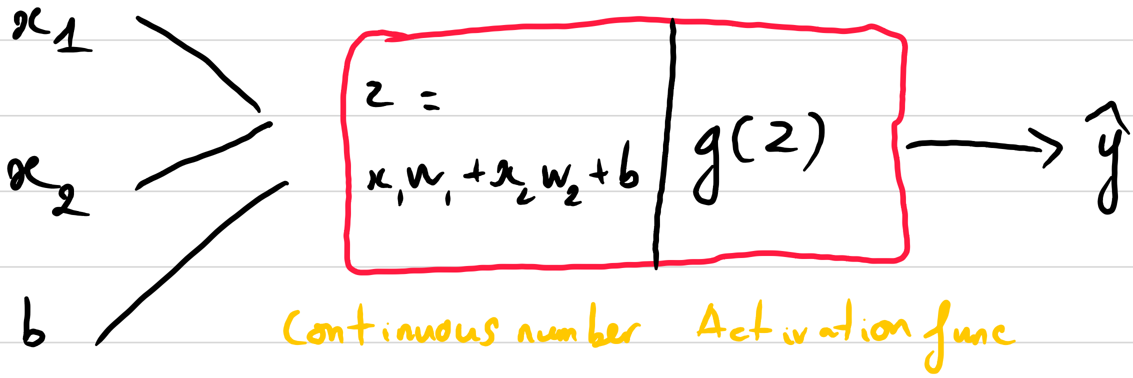
Chain rule

$$\frac{dL}{db} = \frac{dL}{d\hat{y}} \cdot \frac{d\hat{y}}{db}$$

$$\frac{dL}{dw_1} = \frac{dL}{d\hat{y}} \cdot \frac{d\hat{y}}{dw_1}$$

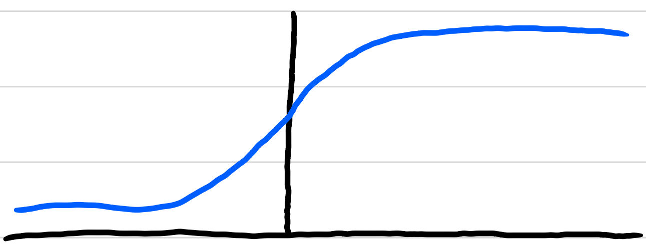
$$\frac{dL}{dw_2} = \frac{dL}{d\hat{y}} \cdot \frac{d\hat{y}}{dw_2}$$

Classification with perceptron



Sigmoid

$$g(z) = \frac{1}{1+e^{-z}}$$



$$\frac{dg}{dz} = g(z) - (1 - g(z))$$

Gradient Descent

Loss:

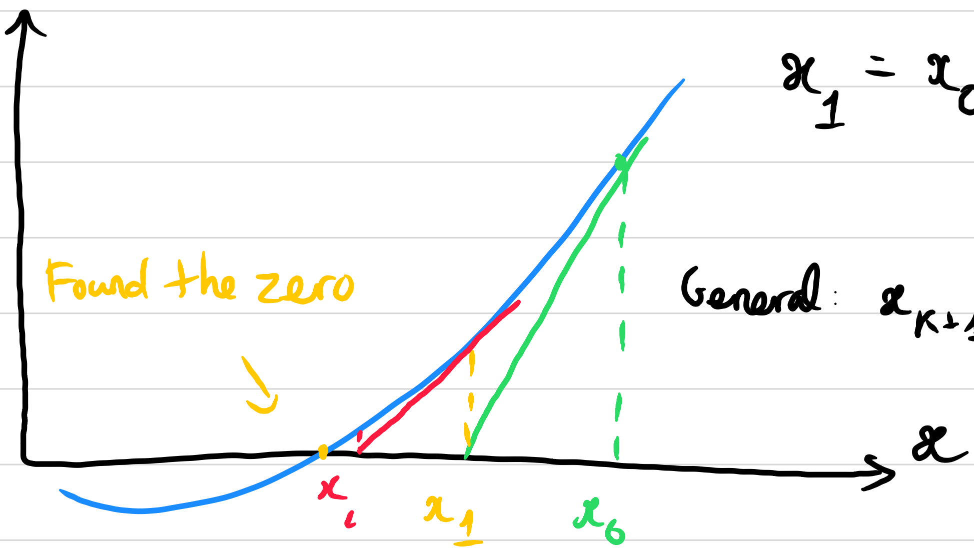
$$L(y, \hat{y}) = -y \ln(\hat{y}) - (1-y)\ln(1-\hat{y})$$

$$w_1 \to w_1 - \alpha \frac{dL}{dw_1}$$

$$w_2 \to w_2 - \alpha \frac{dL}{dw_2}$$

$$b \to b - \alpha \frac{dL}{db}$$

6. Newton's method



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

General: $x_{k+1} = x_k - \dfrac{f(x_k)}{f'(x_k)}$

Newton's method

Goal: find a zero of $f(x)$

1) Start with x_0

2) Update:

$$x_{K+1} = x_K - \frac{f(x_K)}{f'(x_K)}$$

3) Repeat 2 until find the root

NM for optimization

Goal: minimize $g(x) \to$ find zero of $g'(x)$

$f(x) \to g'(x)$ $\quad$ $f'(x) \to (g'(x))'$

1) Start with x_0

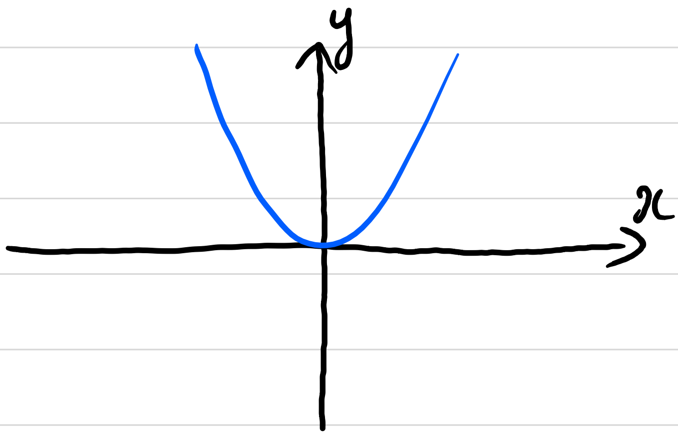
2) Update:

$$x_{K+1} = x_K - \frac{g'(x_K)}{(g'(x_K))'}$$

3) Repeat 2) until find the candidate for minimum

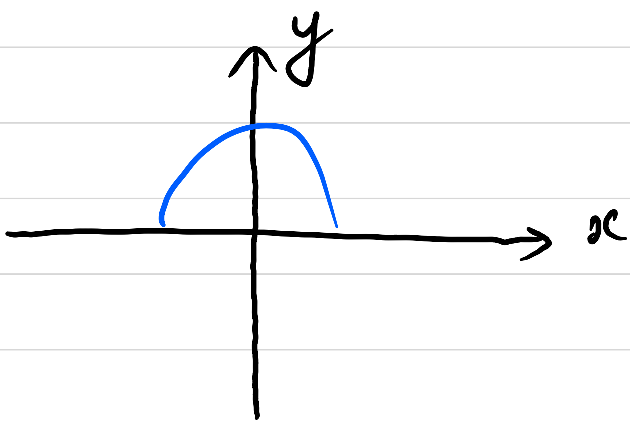
7. Second derivative

<u>Culvature</u>



Concave up or convex

$f''(0) > 0$



Concave down

$f''(0) < 0$

$$\frac{d^2x}{dt^2} > 0 \quad : \quad \text{(local) minimum}$$

$$\frac{d^2x}{dt^2} < 0 \quad : \quad \text{(local) maximum}$$

$$\frac{d^2x}{dt^2} = 0 \quad : \quad \text{Inconclusive}$$

8. Hessian

$$f(x,y) = 2x^2 + 3y^2 - xy$$

$$f(x,y) \begin{cases} \xrightarrow{x} 4x - y \begin{cases} \xrightarrow{x} 4 \\ \xrightarrow{y} -1 \end{cases} \\ \xrightarrow{y} 6y - x \begin{cases} \xrightarrow{x} -1 \\ \xrightarrow{y} 6 \end{cases} \end{cases}$$

$$H = \begin{bmatrix} 4 & -1 \\ -1 & 6 \end{bmatrix}$$

Hessian matrix

9. Newton's method for 2 variables

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} - H^{-1}(x_k, y_k)\, \nabla f(x_k, y_k)$$

$xx \to 2$

$xy \to 0$

$yx \to 0$

$yy \to 6y$