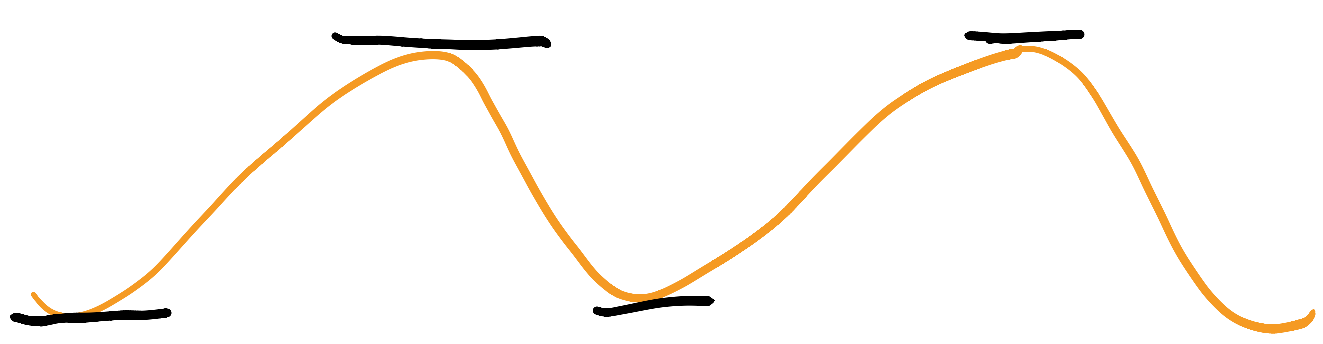
Calculus

- the derivative rate of change between two variables at a particular point

slope : $\frac{\text{rise} \uparrow}{\text{run} \rightarrow}$



- the max / min happen where the derivative = ∅

$y = f(x)$

$f'(x)$
(Lagrange's notation)

$\frac{dy}{dx} = \frac{d}{d(x)} fx$ (Leibinitz's notation)

Common derivatives

Power functions

$f(x) = ax + b$ $\quad f'(x) = a$

$f(x) = ax^2 + bx$ $\quad f'(x) = 2ax + b$

$f(x) = x^{-1}$ $\quad f'(x) = -x^{-2}$

$f(x) = k$ $\quad f'(x) = 0$

↳ where k = constant

Inverse functions $\quad g'(y) = 1/ f'(x)$

trigonometric

$f(x) = \sin(x)$ $\quad f'(x) = \cos(x)$

$f(x) = \cos(x)$ $\quad f'(x) = -\operatorname{sen}(x)$

$f(x) = \tan(x)$ $\quad f'(x) = \sec^2(x) = 1/\cos^2(x)$

Inverse function if a function $f(x)$ does something, $g(x)$ is its inverse if it un-do that thing

$g(f(x)) = x$

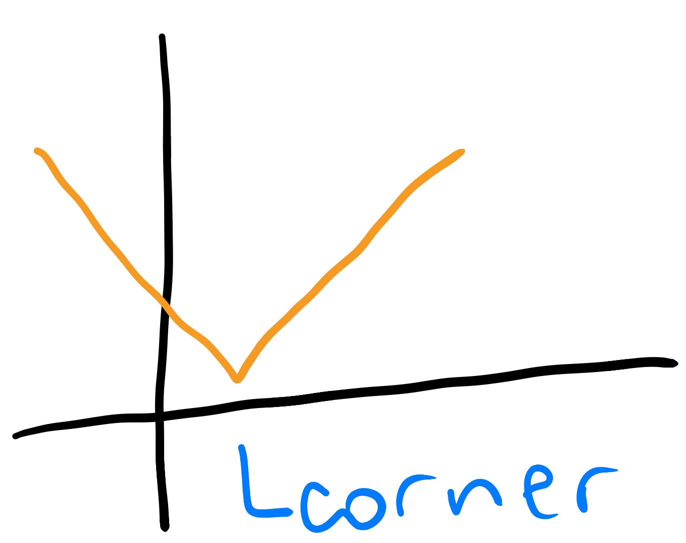
e^x $\qquad f(x) = e^x \qquad f'(x) = e^x$

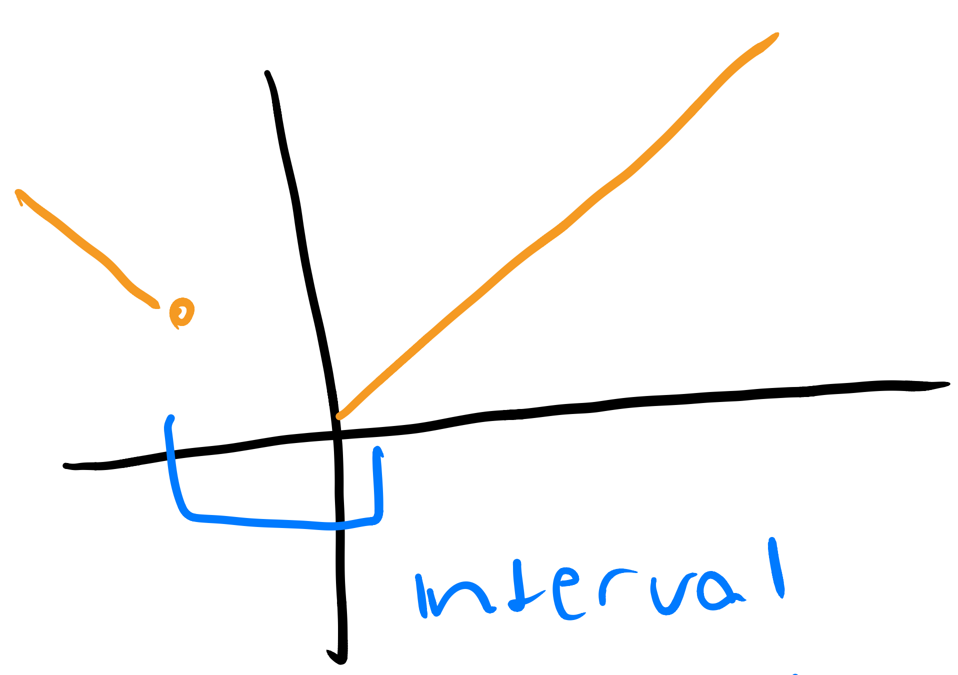
$\log(x)$ $\qquad f(x) = \log(x) \qquad f'(x) =$

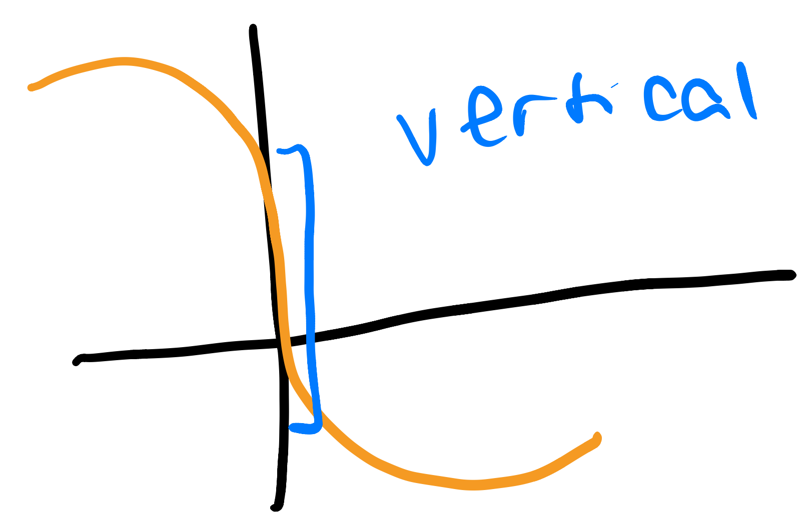
- Non-differentiable functions

Usually if a function has a "corner", it's non-differentiable (the derivative doesn't exist at that corner)

→ if you are drawing a function and you have to lift the pen up, it's non-differentiable







- Rule: Multiplication by scaler

$$k\, f(x) = k\,(f'(x))$$

- Rule: Sum

$$f = g + h \quad \therefore \quad f' = g' + h'$$

- Rule: Product

$$f = g \cdot h \quad \therefore \quad f' = g'h + h'g$$

$e = (1 + 1/n)^n$

$e^x = k \qquad x = \log$

- Rule: Chain rule

$$\frac{d}{dt} f(g(h(t))) =$$

$$\frac{df}{dg} \cdot \frac{dg}{dh} \cdot \frac{dh}{dt} =$$

$$f'(g(h(t))) \cdot g'(h(t)) \cdot h'(t)$$

Partial derivatives

$f(x,y)$

$f_y = \frac{\partial f}{\partial y}$ (partial derivative of f with respect to y)

$f_x = \frac{\partial f}{\partial x}$

(partial derivative of f with respect to x)

1. treat all other variables as constant
2. use regular derivative rules

ie. Given $f(x) = 3x^2y^3$, $\frac{\partial f}{\partial x} = 6xy^3$

Gradients

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

- the gradient is a vector with all the partial derivatives of a function

example $f(x,y) = x^2 + y^2 \therefore \nabla f = \begin{bmatrix} 2x \\ 2y \end{bmatrix}$

---

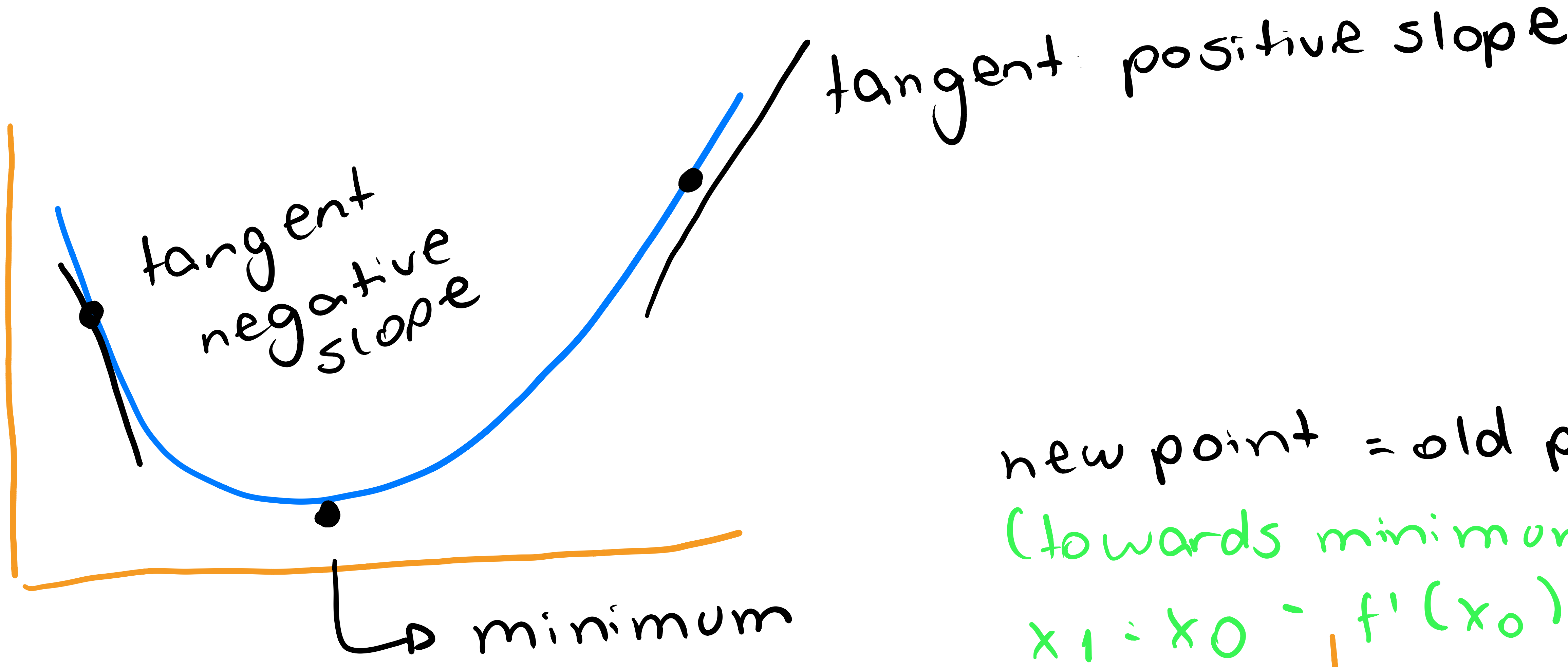
example 2:

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

$$\nabla f(x,y) = \begin{bmatrix} y^2+2 \\ 2xy+3 \end{bmatrix}$$

ps: you treat the other variables as constants.

Gradient descent



new point = old point - slope

(towards minimum)

$x_1 = x_0 - f'(x_0)$

α (learning rate ensures small steps)

Optimisation in neural networks

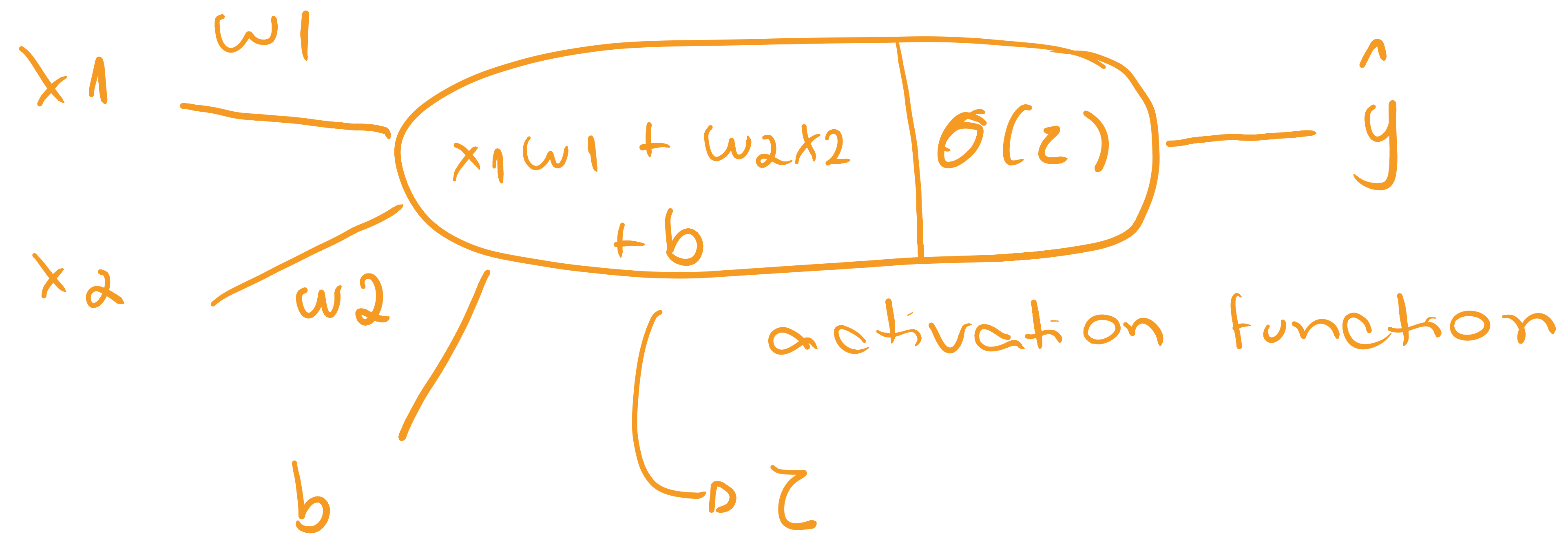
Regression with a perceptron

$$x_1 \underline{w_1} + x_2 \underline{w_2} + \underline{b} = \hat{y}$$

Goal: find best w_1, w_2 and b that minimise loss function

- Loss function: mean squared error

Classification with a perceptron



sigmoid

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

output between 0 and 1 but never 0 or 1

- Loss function: log loss

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

- Newton's method

(Used to find zeros of a function)

$$x_{k+1} = x_k - \frac{f(x_k)}{(x_k)}$$

Second derivative

- Leibniz notation: $\dfrac{d^2 f(x)}{dx^2} = \dfrac{d}{dx}\left(\dfrac{df(x)}{dx}\right)$

- Lagrange notation: $f''(x)$

Understanding the second derivative

X = distance (you are travelling)

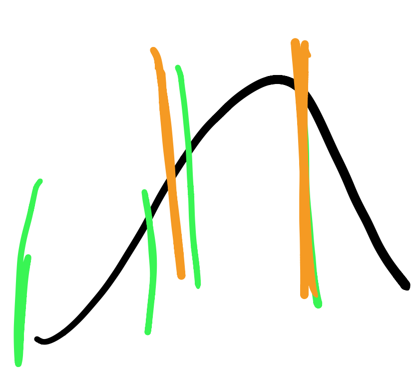
V = velocity you are travelling as a function of time $\dfrac{dx}{dt}$

↳ first derivative

a = acceleration $\dfrac{dv}{dt}$ → second derivative

the second derivative gives a measure of the amount by which a curve deviates from being a straight line.

↳ curvature



concave up (convex)

$\dfrac{d^2x}{dt^2} > 0$ ü

↳ local minimum

concave down

$\dfrac{d^2x}{dt^2} < 0$ ☹

→ local maximum

Hessian Matrix

$$
\begin{tikzcd}
 & & f_{xx}(x,y) \\
 & f_x(x,y) \arrow[ur, "x"] \arrow[r, "y"'] & f_{xy}(x,y) \\
f(x,y) \arrow[ur, "x"] \arrow[dr, "y"'] & & \\
 & f_y(x,y) \arrow[r, "x"] \arrow[dr, "y"'] & f_{yx}(x,y) \\
 & & f_{yy}(x,y)
\end{tikzcd}
$$

1st derivative

2nd derivative

$$
H : \begin{bmatrix} f_{xx}(x,y) & \underline{f_{xy}(x,y)} \\ \underline{f_{yx}(x,y)} & f_{yy}(x,y) \end{bmatrix}
$$

these values are almost always the same