Optimization - 3

-> multi-variable calculus

-> partial derivatives

-> gradients

-> coding : gradient descent.

continuous
but
non-differentiable

$f(x) \rightarrow$ continuous & differentiable

$f'(x) = 0$

ii] $f''(x) > 0$: minima

$f''(x) < 0$: maxima.

price of house $\leftarrow$ (area), (rooms), (location)

x_1 x_2 x_3

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

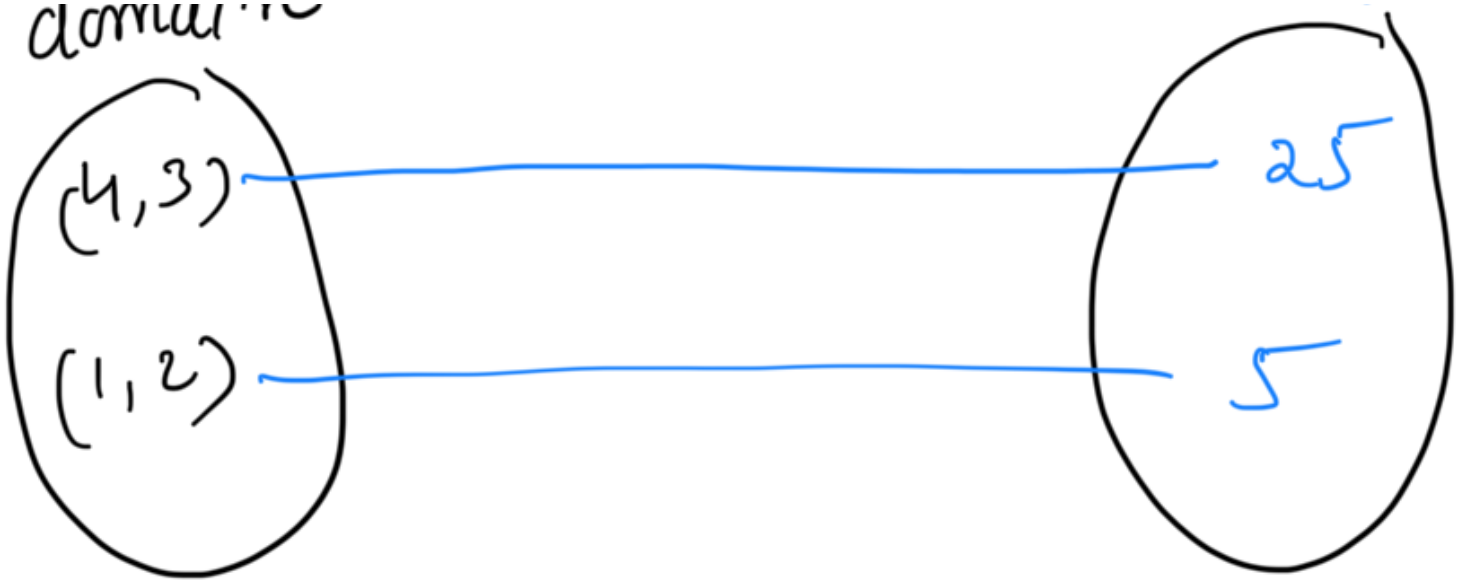
multi-variable function

accept vectors as input and returns a single value as output.

range

domain

domain



Partial Derivative

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

$$\frac{d}{dx} \longrightarrow \frac{\partial}{\partial x}$$

$$\frac{\partial}{\partial x} f(x,y) = \frac{\partial}{\partial x}\left(x^2 + \overset{c}{y^2}\right)$$

$$\frac{\partial}{\partial x} f(x,y) = \ldots$$

$$= 2x + 0$$

$$= 2x$$

$$\frac{\partial}{\partial y} f(x,y) = 0 + 2y$$

$$= 2y$$

$$\left[f(w_1, w_2, w_0) = w_1 x_1 + w_2 x_2 + w_0 \right]$$

$\overline{W}^*$ 2{

$$\boxed{\frac{\partial f}{\ldots} = x_1}$$

$$\ldots = \begin{bmatrix} x_1 \\ \ldots \end{bmatrix}$$

∂w_1

$$\frac{\partial f}{\partial w_2} = x_2$$

$$\frac{\partial f}{\partial w_0} = 1$$

$$\nabla_{\bar{w}} f(w) = \begin{bmatrix} x_2 \\ 1 \end{bmatrix}$$

$$\frac{\partial}{\partial x}(x^1) = 1 \times x^0 = 1 = 1$$

$$\frac{\partial}{dx} x^n = n x^{n-1}$$

$$\boxed{\frac{d}{d\underline{x}}(x) = 1}$$

$$\frac{d}{d\underline{w_0}}(w_0) = 1$$

$$\frac{d}{d\underline{x}}(x+y)$$

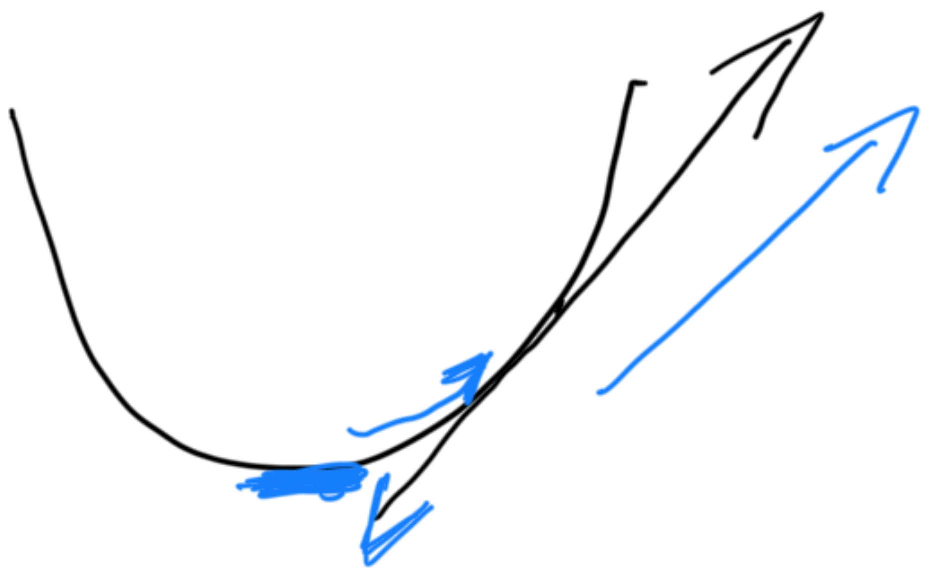
$$= 1 + 0$$

Gradient : vector of partial derivatives.

$$\nabla f(\bar{w}) = \begin{bmatrix} \frac{\partial}{\partial w_1} f \\ \frac{\partial}{\partial w_2} f \\ \vdots \\ \frac{\partial}{\partial w_0} f \end{bmatrix}$$

Gradient.

first principle



magnitude

vector → direction → of steepest ascend

Gradient ascent

$$f(x,y) = 3 \cdot \log(xy) + 4y^2x^3$$

$$\frac{\partial f}{\partial x} = 3 \cdot \frac{1}{xy} \frac{\partial}{\partial x}(xy) + 4y^2 \times 3x^2$$

$$= \frac{3}{x\cancel{y}} \times \cancel{y} + 12y^2x^2$$

$$= \frac{3}{x} + 12y^2x^2$$

$$\frac{\partial}{\partial y}(f) = \frac{3}{y} + 4x^3 \times 2y$$

$$= \frac{3}{y} + 8x^3y$$

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

$f = x^2 + y^2$

$$\nabla f = \begin{bmatrix} 2x \\ 2y \end{bmatrix} = \begin{bmatrix} 6 \\ 8 \end{bmatrix}$$

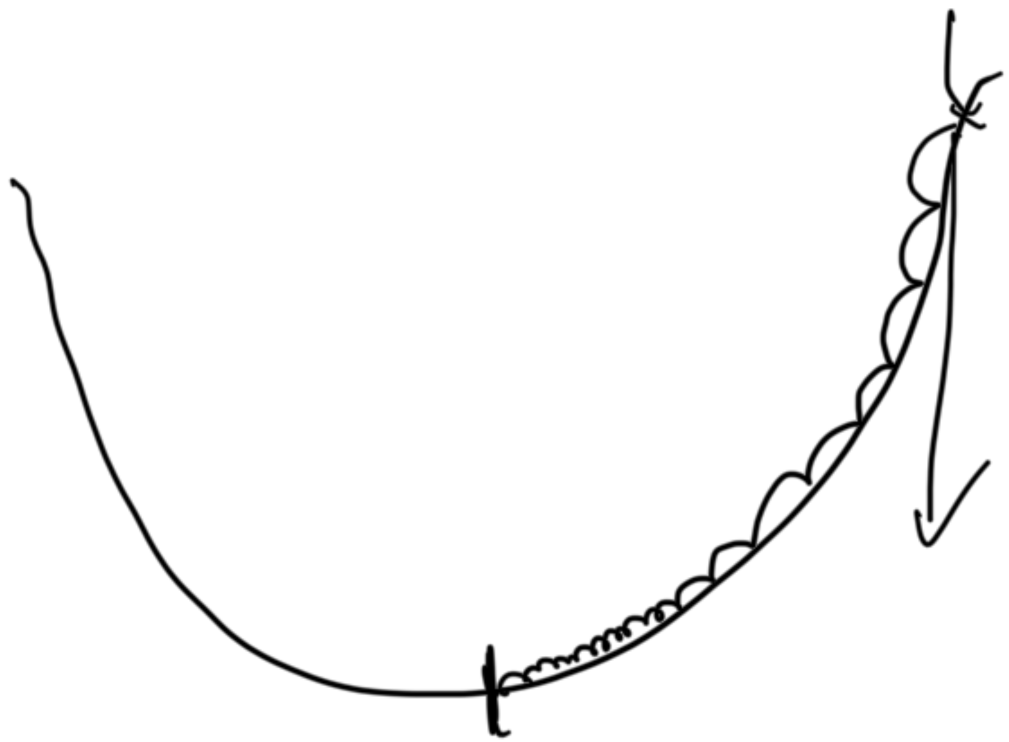
$3, 4 \quad =$

$x^{(0)}$

learning rate
$[0.1 - 0.5]$

$$\boxed{x^{(1)} = x^{(0)} - \eta \frac{\partial}{\partial x}(f)}$$

x^2 $\qquad (x') - \eta \frac{\partial}{\partial x}(f)$



Generalised G·D

$$\eta \overbrace{\left[\nabla \alpha \left(X, \bar{w}, w_0\right)\right.}^{0}$$

$$\bar{W}^{(t+1)} = W - \eta \boxed{\nabla \mathcal{L}(\;\;)}$$

w_1

w_2

$$w_3 = w_3 - \eta \frac{\partial}{\partial w_3} \mathcal{L}(x, \bar{w}, w_0)$$