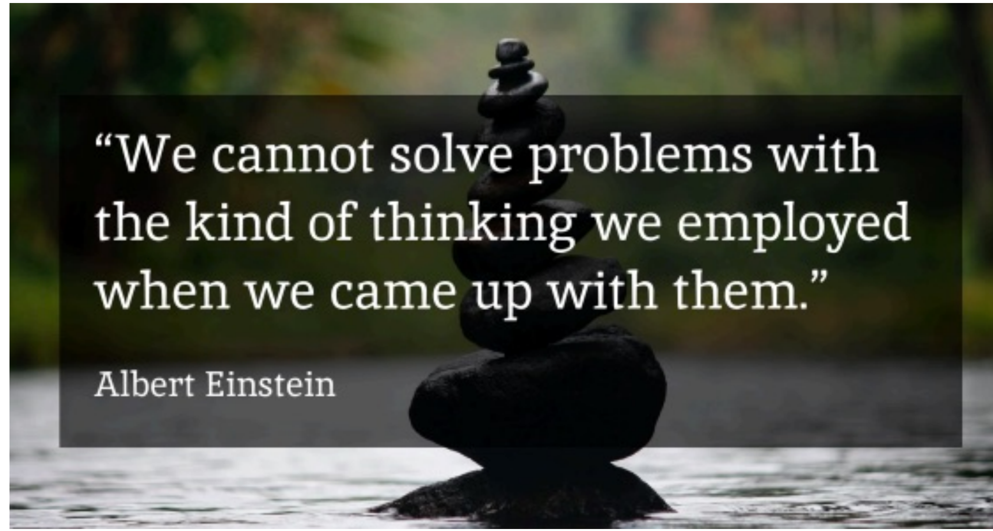
Gradient Descent



Agenda :

→ Multivariate Calculus

→ Partial Derivatives

→ Gradients

Recap

Derivatives : $f(x) \to$ Continous & differentiable

$$\frac{d\,f(x)}{dx} = f'(x) = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

Maxima/minima : To find Candidate Points for maxima & minima.

find x, wh: $f'(x) = 0$

if $f''(x) > 0$: minima

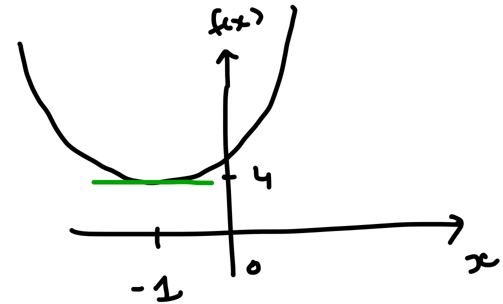
$f''(x) < 0$: maxima

Rules:

1. Linearity
2. Product
3. Quotient
4. Chain

Q: $f(x) = (x+1)^2 + 4$. Minima w/r $x = ?$

$$x^* = \underset{x}{\operatorname{argmin}}\ f(x)$$



$$x^* = \underset{x}{\operatorname{argmin}}\ (x+1)^2 + 4$$

$\text{minimum}(f(x)) = 4$

at what x does minimum happen: -1

"minima"

$$\frac{df(x)}{dx} = \frac{d(x+1)^2 + 4}{dx}$$

$$\frac{df(x)}{dx} = 2(x+1)$$

$$\frac{df(x)}{dx} = 0$$

$$2(x+1) = 0$$

$$\boxed{x = -1}$$

How can we determine if a critical point is a minimum or maximum for a given function?

38 users have participated

A	By observing the function's behavior at that point.	0%
B	By checking if the function value is positive or negative at the critical point.	5%
C	By finding the point where the first derivative is zero.	34%
D ✓	By evaluating the second derivative at the critical point	61%

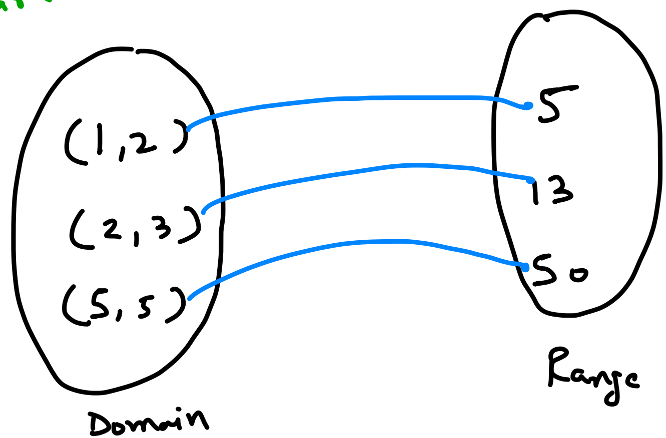
CC Multivariate Calculus

"Multivariate functions"

$$f(\underbrace{x, y}_{\text{multiple input var.}}) = x^2 + y^2$$

input → $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$

input → $[1, 2]^T$



input → vector of all independent var.

output

Scalar (Real No.)

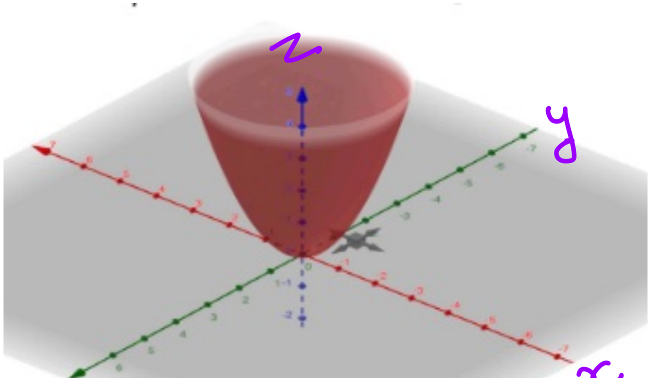
III Partial Derivatives

$f(x) = x^2$

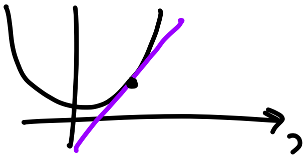
$$z = f(x, y) = x^2 + y^2$$

input

$$\begin{bmatrix} x \\ y \end{bmatrix}$$



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



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

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

gradient

$$\boxed{\nabla_{x,y} f(x,y)} = \begin{bmatrix} \dfrac{\partial f(x,y)}{\partial x} \\ \dfrac{\partial f(x,y)}{\partial y} \end{bmatrix}$$

Another Example

$$f(w_1, w_2, w_0) = w_1 x_1 + w_2 x_2 + w_0$$

$$\nabla_{w,w_0} f(w, w_0) = \begin{bmatrix} \frac{\partial f}{\partial w_1} \\ \frac{\partial f}{\partial w_2} \\ \frac{\partial f}{\partial w_0} \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix}$$

$w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$

$$\frac{\partial (w_1 x_1 + \cancel{w_2 x_2} + \cancel{w_0})}{\partial w_1} = x_1 \frac{\partial w_1}{\partial w_1} = x_1$$

$$\frac{\partial (\cancel{w_1 x_1} + w_2 x_2 + \cancel{w_0})}{\partial w_2} = x_2 \cdot \frac{\partial w_2}{\partial w_2} = x_2$$

gradient Vector of f w.r.t w & w_0 $\nabla_{w,w_0} f$

$[w_1, w_2]$

$w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$

Single variable → Derivatives

Multiple variable → Partial Derivatives = gradients

$$f(x,y) = 2x^2 + 4xy$$

$$\frac{\partial f}{\partial x} = 4x + 4y = 4(x+y)$$

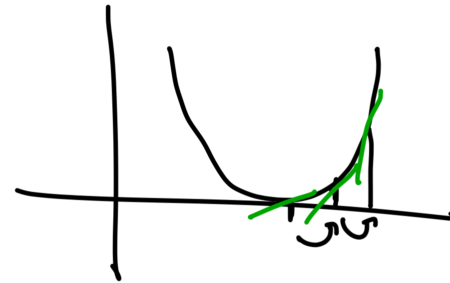
$$\frac{\partial f}{\partial y} = 4x$$

$$(3,4) = \begin{bmatrix} 28 \\ 12 \end{bmatrix}$$

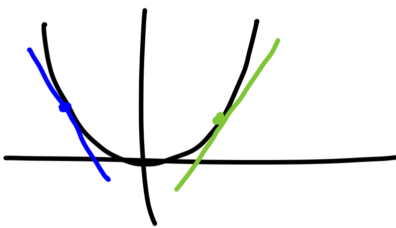
$$f = x^2 + y^2$$

$$\begin{bmatrix} 2x \\ 2y \end{bmatrix}$$

$(x,y) = (1,1)$



$f(x) = x^2$



Q $f(\underbrace{x,y}_{\bar{x}}) = 3\log(xy) + 4y^2x^3$

$\bar{x} = \begin{bmatrix} x \\ y \end{bmatrix}$

$$\nabla_{\bar{x}} f(\bar{x}) = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} = \frac{\partial \left[3\log(xy) + 4y^2x^3\right]}{\partial x}$$

$$= 3\frac{\partial \log xy}{\partial xy} \cdot \frac{\partial xy}{\partial x} + 4y^2\frac{\partial x^3}{\partial x}$$

$$= 3 \cdot \frac{1}{x\cancel{y}} \cdot \cancel{y} + 12y^2x^2$$

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

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

$$= 3\frac{\partial \log xy}{\partial y} + 4x^3\frac{\partial y^2}{\partial y}$$

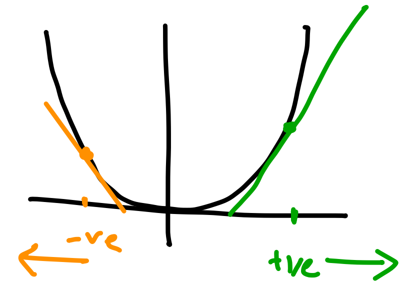
$$= 3 \cdot \frac{\partial \log xy}{\partial xy} \cdot \frac{\partial xy}{\partial y} + 8x^3y$$

$\vdots$

$\vdots$

$$= 3 \cdot \frac{1}{\cancel{x}y} \cdot \cancel{x} + 8x^3y$$

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



derivative

Partial derivatives = gradient = dir^n of steepest ascend

III Gradient Descent

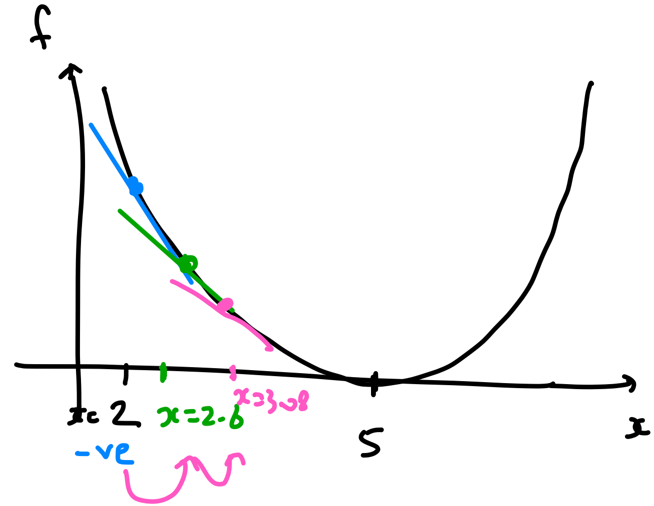
$$f(x) = (x-5)^2$$

$$x^* = \underset{x}{\operatorname{argmin}} (x-5)^2$$

Best value of x

$\Downarrow$

Gradient Descent



① randomly init variable. (x)

② repeate until Convergence {

$$x = x - \eta \cdot \frac{\partial f}{\partial x}$$

}

$\eta \geq 0$ = learning-rate : step-size

$$= 2 + \eta \cdot 6$$

$$\hookrightarrow 0.1$$

$$x_{update} = 2 + 0.6$$

$$= 2.6$$

$$2(x-5)$$

$$x = x - \eta \boxed{\frac{\partial f}{\partial x}}$$

$$= 2.6 + 0.1(4.8)$$

$$= 2.6 + 0.48$$

$$= 3.08$$

$$x = x - \eta \cdot \frac{\partial f}{\partial x}$$

$$= 10 - 0.1 \times 10$$

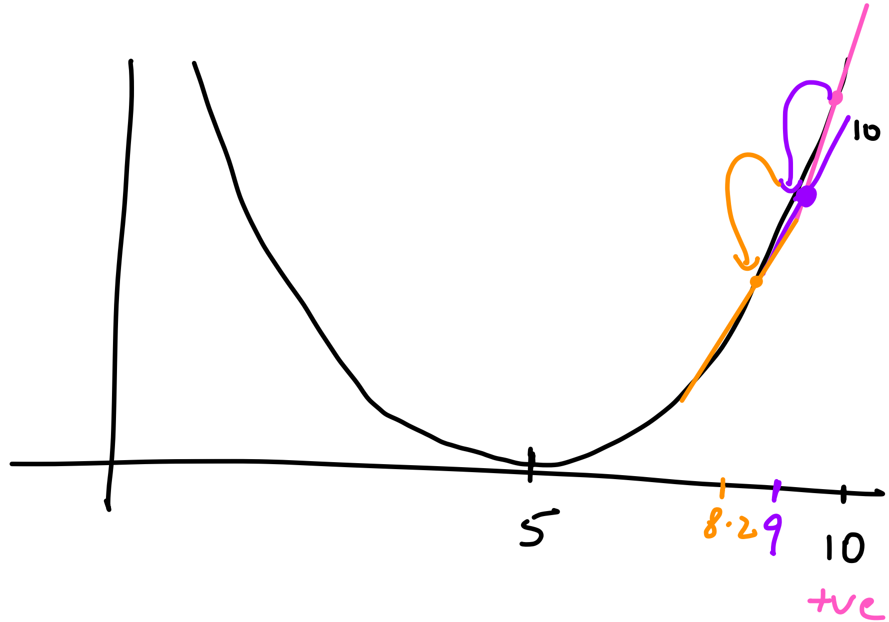
$$= 10 - 1$$

$$\boxed{x_{updated} = 9}$$

$$= 9 - 0.1 \times 8$$

$$= 9 - 0.8$$

$$= 8.2$$

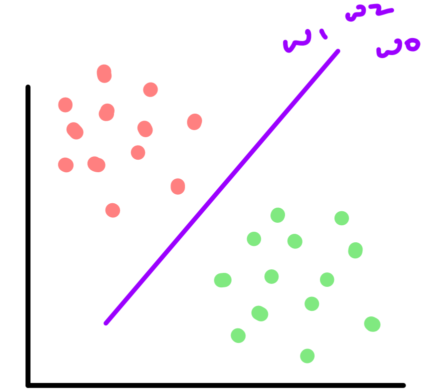


$2(x-5)$

$2(x-5)$

$= 8$

$$L(\bar{w}, w_0) = -\frac{1}{n}\sum_{i=1}^{n = \#\,\text{data pts}} \left(\frac{\bar{w}^T x^i + w_0}{\|w\|}\right)\cdot y^i$$



$$w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$w = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix}$$

$$w^*, w_0 = \underset{w, w}{\arg\min}\ L(w, w)$$

1) Randomly init all vars $\bar{w}$, w_0

2) repeat until convergence {

$$\bar{w} = \bar{w} - \eta \cdot \frac{\partial L}{\partial \bar{w}}$$

$$w_0 = w_0 - \eta \cdot \frac{\partial L}{\partial w_0}$$

}

$$\bar{w} = \bar{w} - \eta \cdot \frac{\partial L}{\partial \bar{w}}$$

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} - \eta \cdot \begin{bmatrix} \frac{\partial L}{\partial w_1} \\ \frac{\partial L}{\partial w_2} \\ \vdots \end{bmatrix}$$

$$\left\{ \begin{aligned} w_1 &= w_1 - \eta \cdot \frac{\partial L}{\partial w_1} \\ w_2 &= w_2 - \eta \cdot \frac{\partial L}{\partial w_2} \\ w_0 &= w_0 - \eta \cdot \frac{\partial L}{\partial w_0} \end{aligned} \right\}$$

$$\nabla_{\bar{w}} L$$