Agenda

- Intro to Multivariate Calculus
- Partial Derivatives
- Gradients
  - Representation of Gradients
  - Geometric Interpretation
- Computing Gradients of Any Function
  - Rule Based
  - Ab initio Code
- Intuition of Gradient Descent
- Generalization of Gradient Descent

Recap

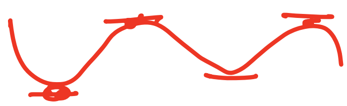
① Derivative : $f(x) \Rightarrow$ Continuous & differentiable

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

② CP $\Rightarrow$ $f'(x) = 0$

Maxima $f''(x) < 0$ $\qquad x$

Minima $f''(x) > 0$

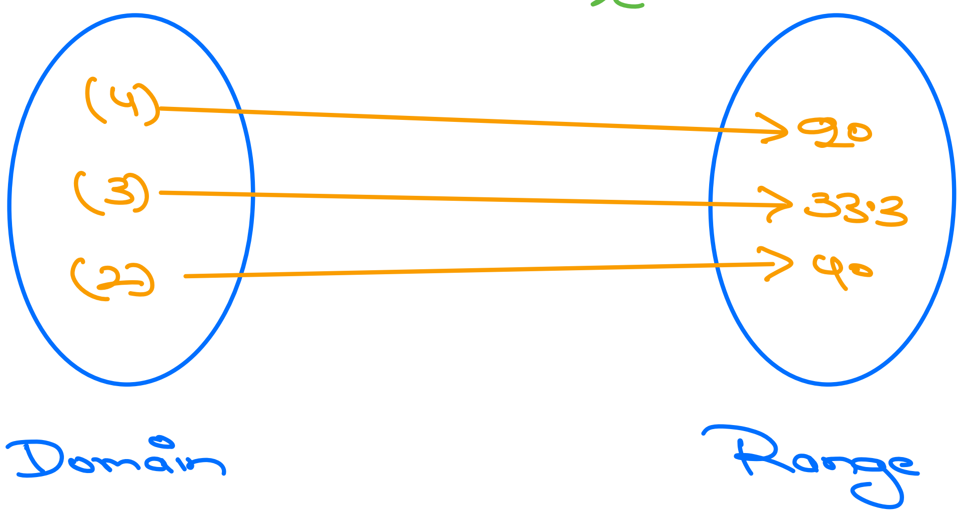


CP

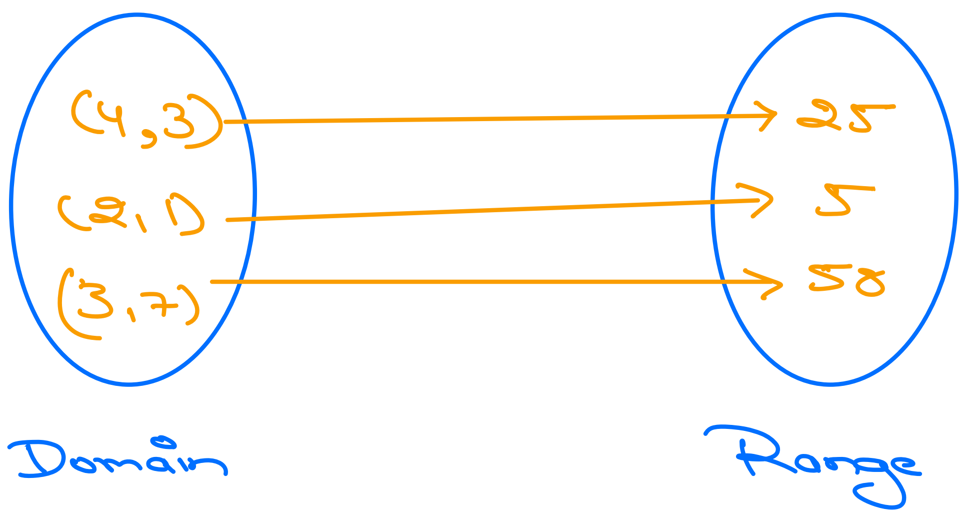
Intro to Multivariate Calculus

Univariate: $f(x) = 3x + \log x$

$f'(x) = 3 + \frac{1}{x}$

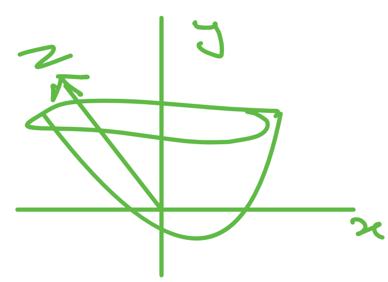


Multivariate: $f(x, y) = x^2 + y^2$

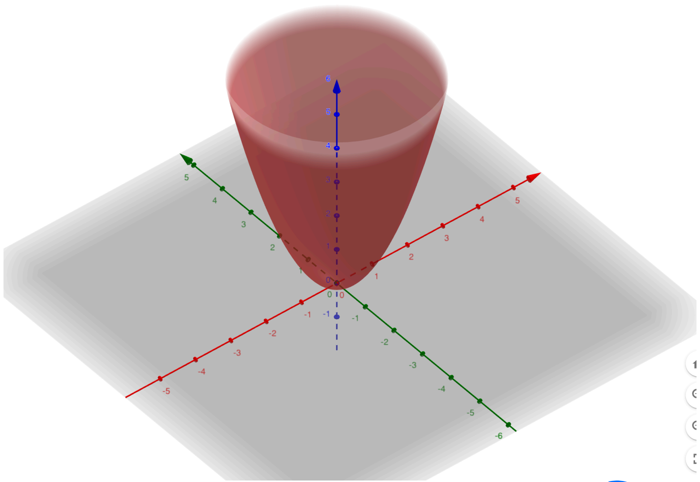


- A mapping that takes multiple Vars as input and Return single Value as output

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



Partial Derivatives



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

Partial Derivative w.r.t Each Variable

"While calculating partial Derivative w.r.t One Variable treat all Other Vars as Constant"

$$z = x^2 + y^2$$

$$\frac{\partial (x^2)}{\partial x}$$

$$\frac{\partial c}{\partial x} = 0$$

① $\frac{\partial z}{\partial x} \Rightarrow 2x + 0$

② $\frac{\partial z}{\partial y} \Rightarrow 0 + 2y$

Q Given Function

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

① $$\frac{\partial f(w_1, w_2, w_0)}{\partial w_1} = x_1 + 0 + 0 \quad \boxed{\frac{\partial x_2}{\partial w_1}}$$

② $$\frac{\partial f(w_1, w_2, w_0)}{\partial w_2} = x_2$$

③ $$\frac{\partial f(w_1, w_2, w_0)}{\partial w_0} = 0 + 0 + 1$$

Representation of Gradients

The partial derivative we calculated in previous step can be combined and represented as a single vector called as Gradient

$$\nabla_{\bar{w}} f(\bar{w}) = \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix}$$

$$\downarrow$$

$$\begin{bmatrix} \frac{\partial f}{\partial w_1} \\ \frac{\partial f}{\partial w_2} \\ \frac{\partial f}{\partial w_0} \end{bmatrix}$$

* Derivative : Single Var $\quad f(x) \rightarrow f'(x)$

* Gradient : Multiple Var $\quad f(x, y) = f'_x$
f'_y

Computing Gradients of Any Function

1. Rule Based

Q1 $f(x,y) = 3\log(x \cdot y) + 4y^2x^2$

$$\nabla_{\bar{x}} f(\bar{x}) \Rightarrow \begin{bmatrix} \dfrac{\partial f(\bar{x})}{\partial x} \\ \dfrac{\partial f(\bar{x})}{\partial y} \end{bmatrix} \qquad \bar{x} = \begin{bmatrix} x \\ y \end{bmatrix}$$

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

$$3\frac{1}{xy} \times y + 8y^2x$$

$$f(g(x))$$

$$f'(g(x)) \times g'(x)$$

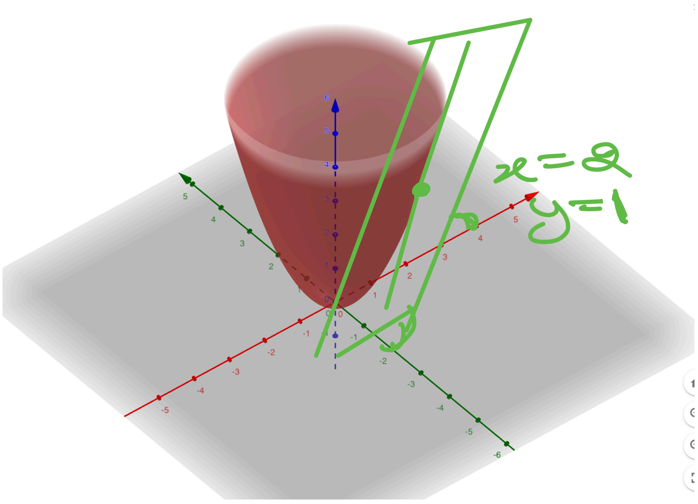
$$\frac{\partial\, xy}{\partial x} = y$$

$$\frac{\partial f}{\partial x} = \frac{3}{x} + 8y^2x$$

$$\nabla_{\bar{x}} f(\bar{x}) \Rightarrow \begin{bmatrix} \dfrac{3}{x} + 8y^2x \\ \dfrac{3}{y} + 8x^2y \end{bmatrix}$$

$\underline{Q}$ 2 $\quad f(x_1, x_2) = \log(x_1) + \frac{x_1}{x_2} e^{x_1 + x_2}$

H.W $\quad \nabla_{\bar{x}} f(x, y) = ?$



$\nabla_x \Rightarrow$ rate of change of x

rate of change of y

rate of change in multiple Direction

② Ab initio Code

① $\quad f(x) = x^2$

$$f'(x) \Rightarrow \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

③ $\quad f(x, y) = x^2 + y^2$

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

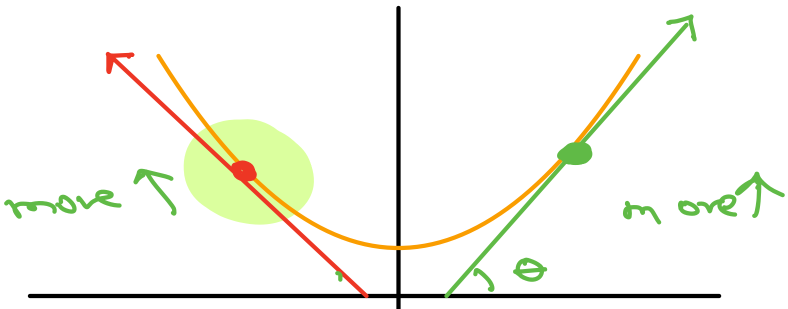
$$\frac{\partial f(x,y)}{\partial y} \Rightarrow \lim_{\Delta y \to 0} \frac{f(x, y + \Delta y) - f(x, y)}{\Delta y}$$

x 1 2 3 4

y 10 20 30 40

$$\begin{bmatrix} (1,10) & (2,10) & (3,10) & (4,10) \\ (1,20) & (2,20) & (3,20) & (4,20) \\ (1,30) & 2 & \dots & \end{bmatrix}$$

Intuition of Gradient Descent



If we move in direction of slope the value of $f(x)$ increases

Goal of Gradient Descent
⇒ Minimize Given function and finding out parameters

Steps of Gradient Descent :

Step 1 : Initialize var/parameter randomly

Step 2 : Find the Slope of tangent at the initial point

Step 3 : Update the initial Values using GD Rule

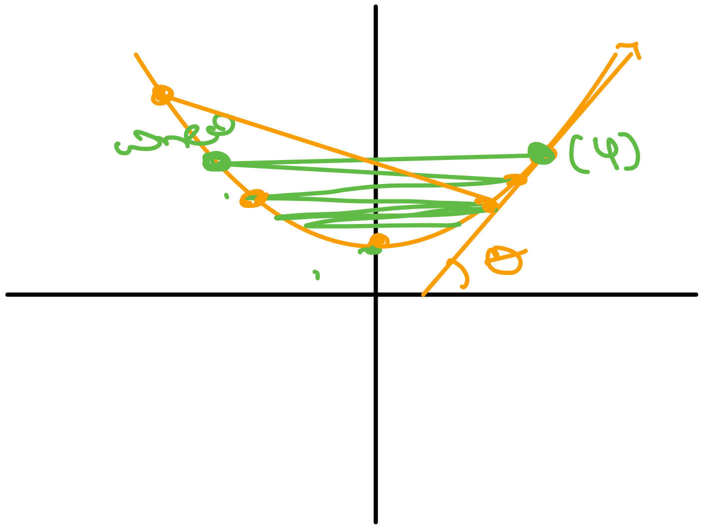
Step 4 : Repeat

$$w_1 = w_1^{current} - \eta \frac{\partial f(w)}{\partial w}$$

Hyperparameter ⇒ learning_rate

$$w_1^{new} = w_1^{current} - 0.01 \times \frac{\partial f(w)}{\partial w_1}$$

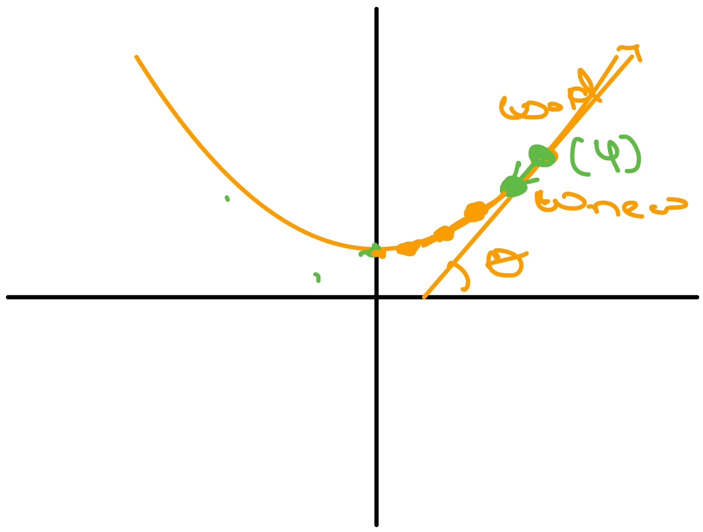
$$w_1 - \frac{\partial f(x)}{d w}$$



slope = $\tan\theta$

$\downarrow$

0, ∞



$\tan\theta = 8$

0.01×8

$\Rightarrow 0.08$

Generalization of Gradient Descent

$$G(X, \bar{w}, w_0) = \sum \frac{(\bar{w}^T x + w_0) * y_i}{\|w\|}$$

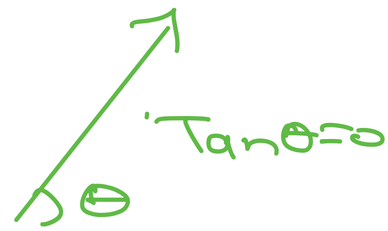
$$L(X, \bar{w}, w_0) = -\sum \frac{(\bar{w}^T x + w_0) * y_i}{\|w\|}$$

$$\nabla_{(\bar{w})} \Rightarrow \bar{w}^{(t)} - \eta \nabla_{\bar{w}} L(x, \bar{w}, w_0)$$

$$w_0 \Rightarrow w_0 - \eta \nabla_{w_0} L(x, \bar{w}, w_0)$$

$$\boxed{w = w + \eta \frac{\partial L}{\partial w}}$$

Gradient Ascent



Descent ↓

Ascent ↑