AI VIETNAM
AI Course 2023

Derivative and its Application

Vinh Dinh Nguyen
PhD in Computer Science

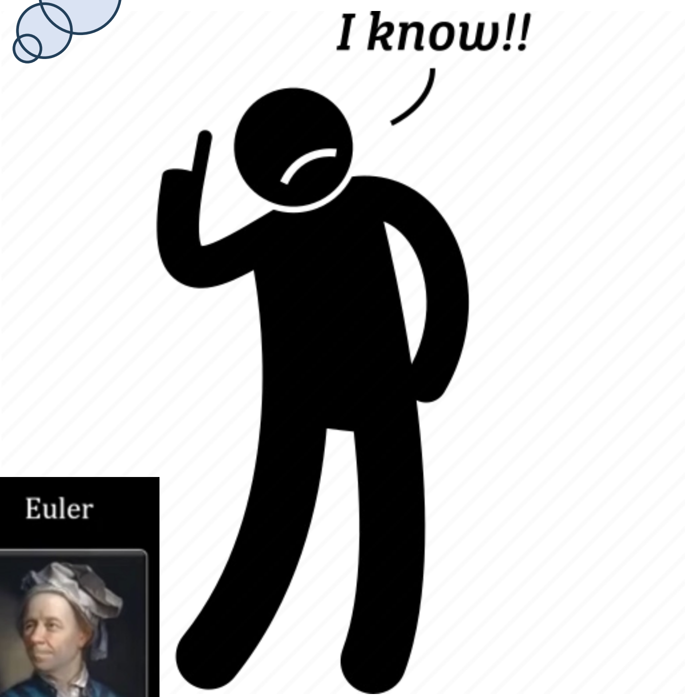
Year 2023

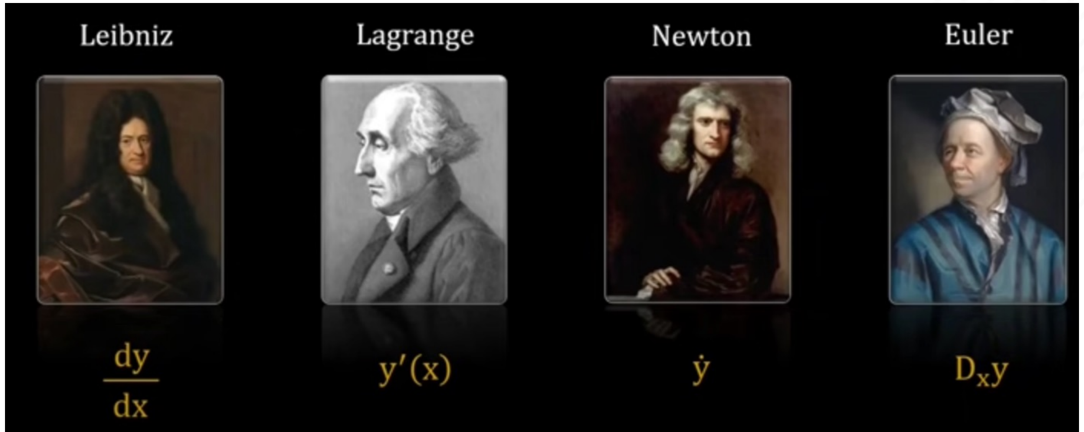

WHAT
HAVE YOU
LEARNED

Đạo hàm rất quan trọng trong ML. Mình vẫn chưa hiểu đạo hàm là gì?

Để hiểu đạo hàm bẩn cần biết:

- Limits of function (giới hạn hàm số)
- Slope of the function (độ đốc hàm số)
- Secant line (đường cát tuyến)
- Targent line (đường tiếp tuyến)

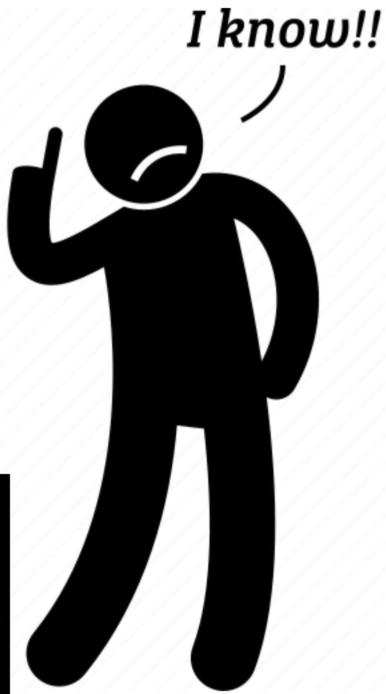


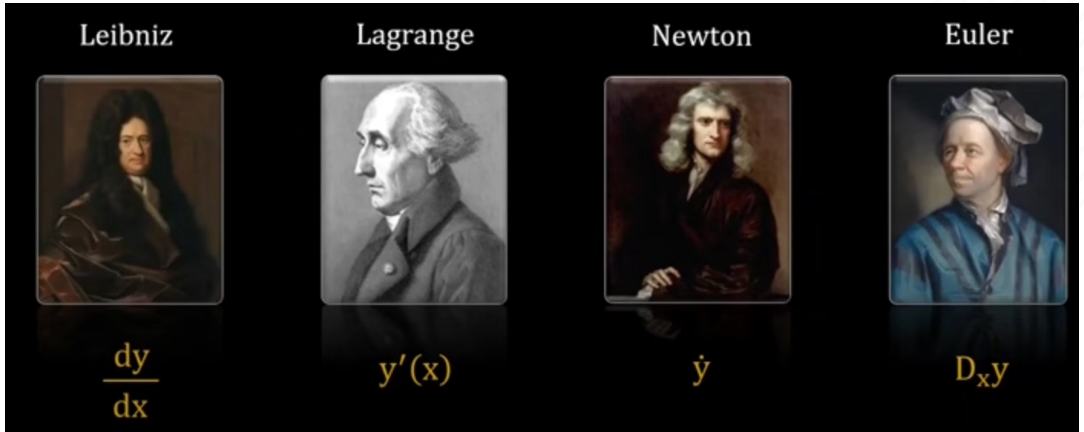


Đạo hàm rất quan trọng trong ML. Mình vẫn chưa hiểu đạo hàm là gì?

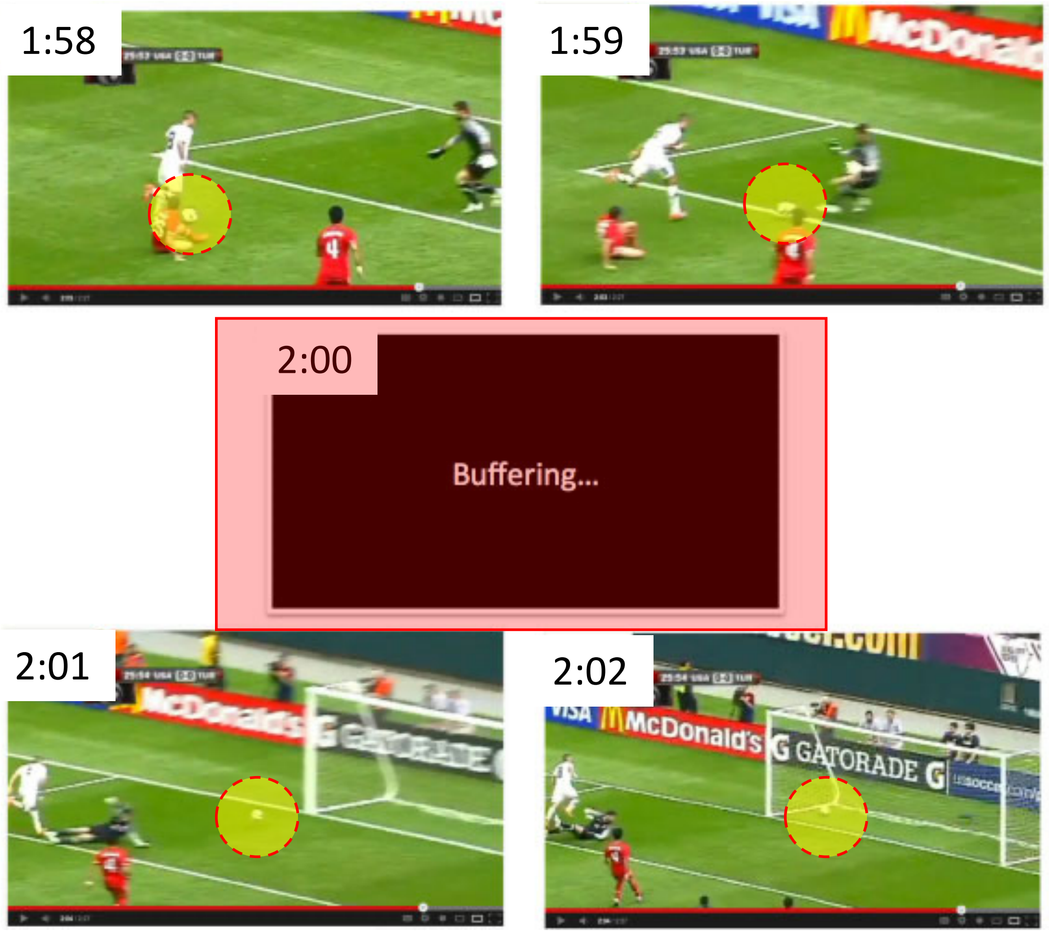
Để hiểu đạo hàm bần cần biết:

- Limits of function (giới hạn hàm số)
- Slope of the function (độ dốc hàm số)
- Secant line (đường cát tuyến)
- Targent line (đường tiếp tuyến)





Limit (Behind the Scene)

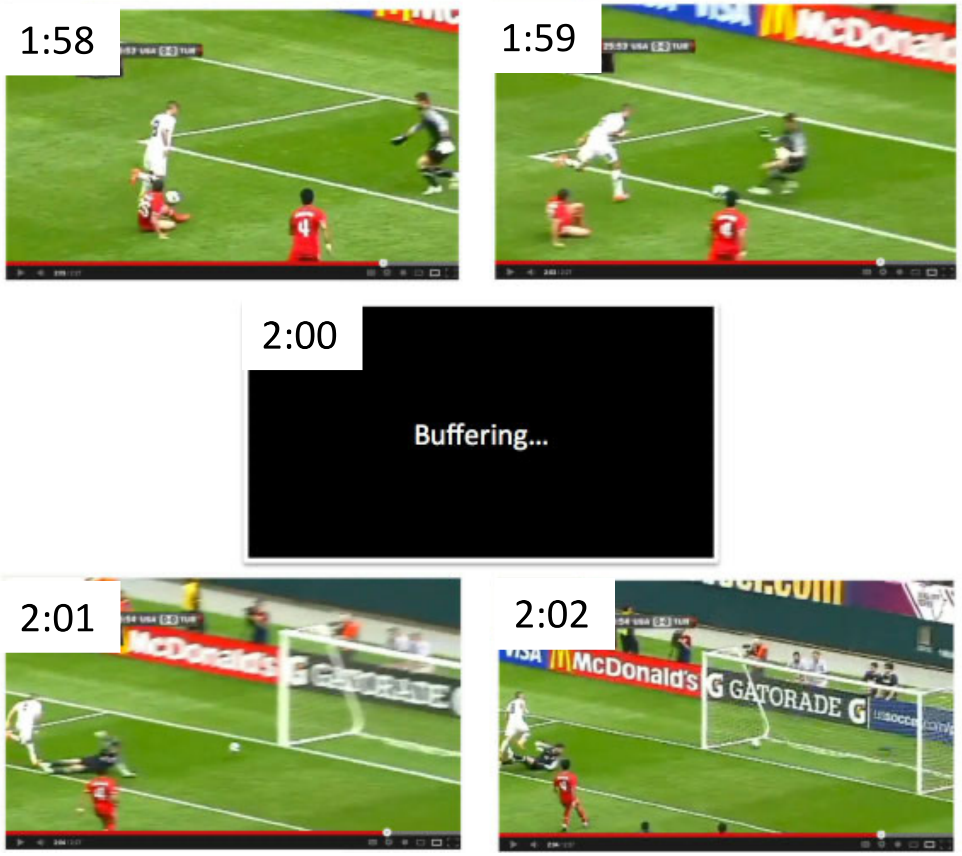


Chào bạn, tôi là Limits, giang hồ hay gọi tôi là "Giới hạn"

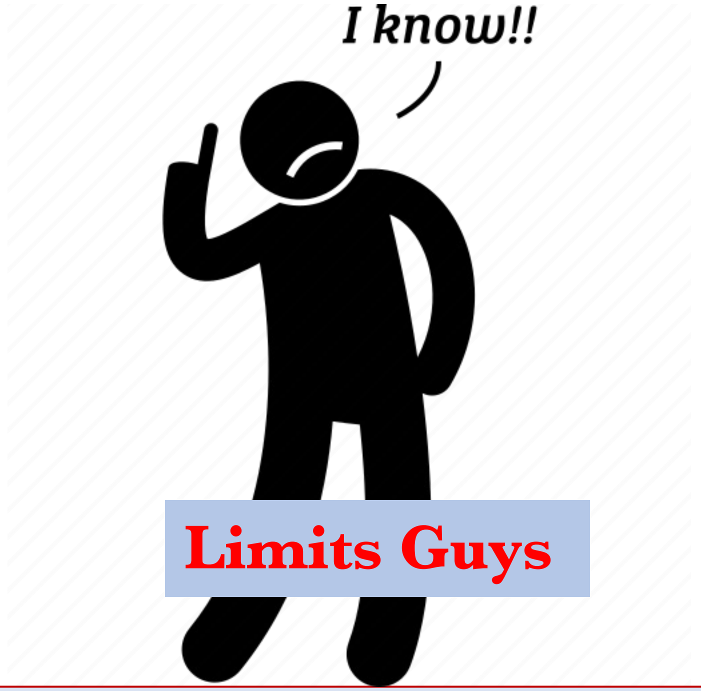
Giả sử bạn đang xem một trận bóng đá, nhưng do đường truyền không ổn định, và kết quả là các bạn không biết chuyện gì xảy ra ở phút thứ 2:00

Limit (Behind the Scene)

Điều gì đang diễn ra ở phút thứ 2:00 nhỉ?



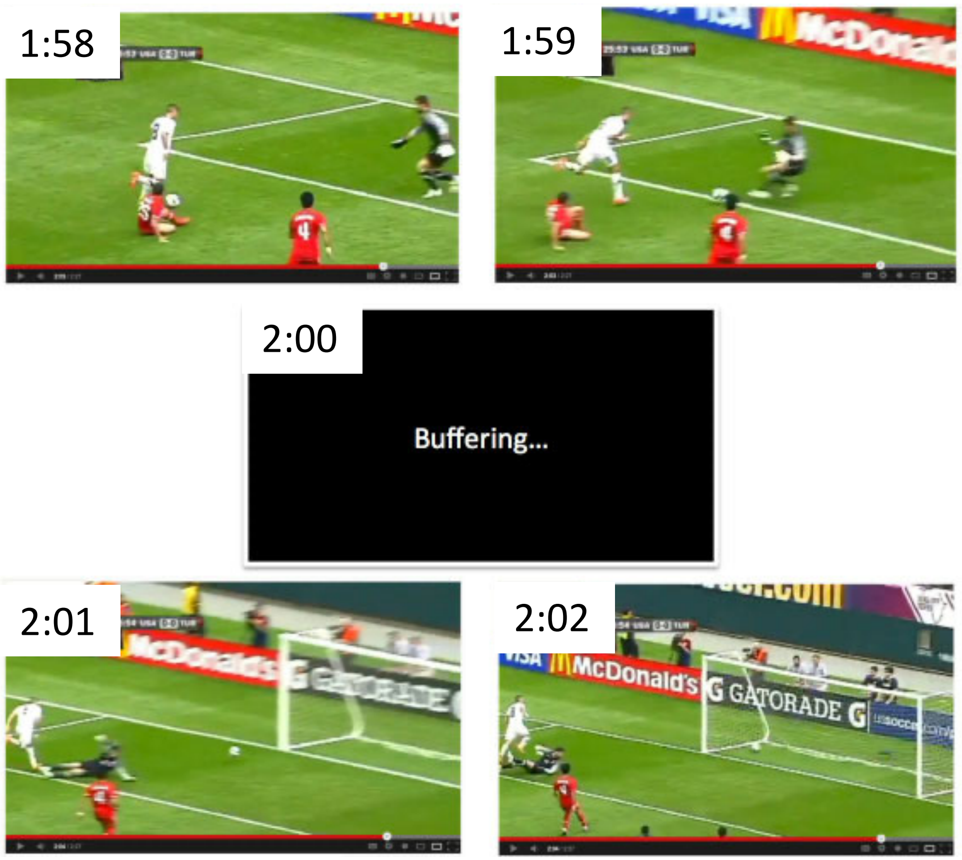
Quá Dễ. Chỉ cần lấy hai thời gian liền kề (1:59 và 2:01) và dự đoán quả bóng sẽ ở đâu đó ở giữa thời gian này



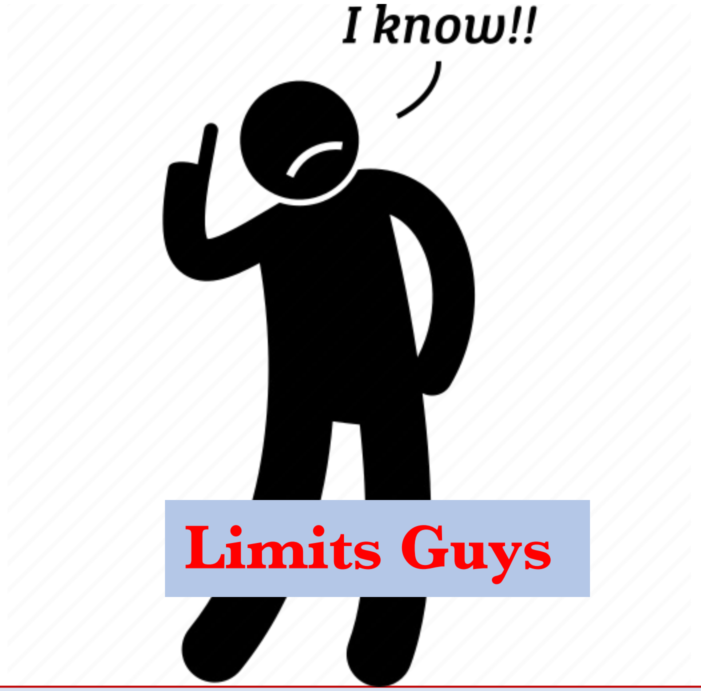
Giả sử bạn đang xem một trận bóng đá, nhưng do đường truyền không ổn định, và kết quả là các bạn không biết chuyện gì xảy ra ở phút thứ 2:00

Limit (Behind the Scene)

Điều gì đang diễn ra ở phút thứ 2:00 nhỉ?



Dự đoán của mình là "Vào lúc 2:00, quả bóng nằm giữa vị trí lúc 1:59 và 2:01". Bạn thấy sao?

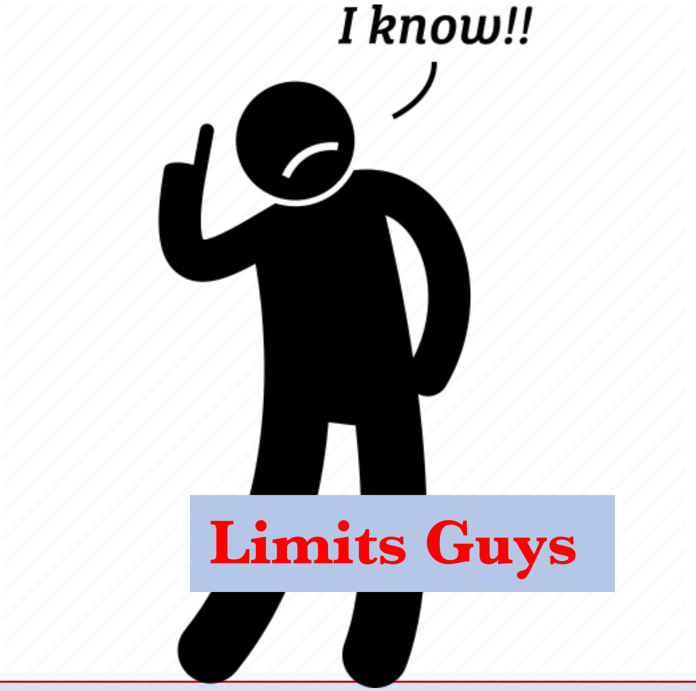


Giả sử bạn đang xem một trận bóng đá, nhưng do đường truyền không ổn định, và kết quả là các bạn không biết chuyện gì xảy ra ở phút thứ 2:00

Limit (Behind the Scene)

Điều gì đang diễn ra ở phút thứ 2:00 nhỉ?

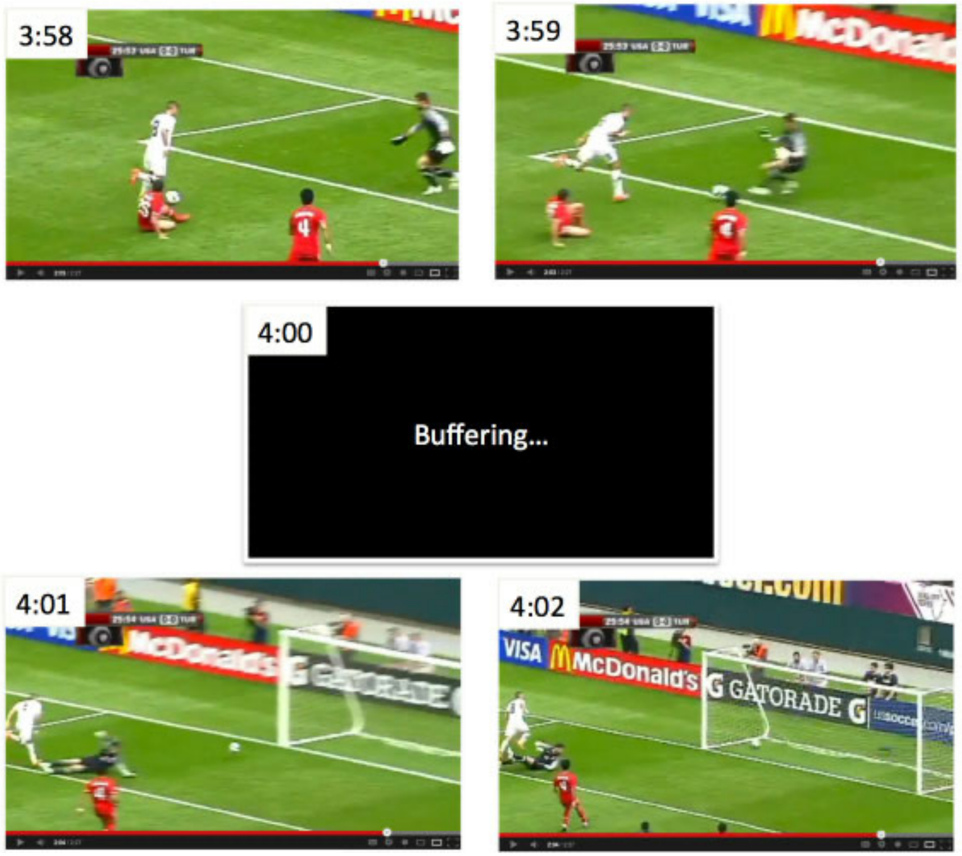
Với một máy quay chuyển động chậm, thậm chí mình có có thể nói “Lúc phút 2:00, quả bóng nằm tại vị trí của nó giữa phút 1:59.999 và phút 2:00.001”.



Giả sử bạn đang xem một trận bóng đá, nhưng do đường truyền không ổn định, và kết quả là các bạn không biết chuyện gì xảy ra ở phút thứ 2:00

Limit (Behind the Scene)

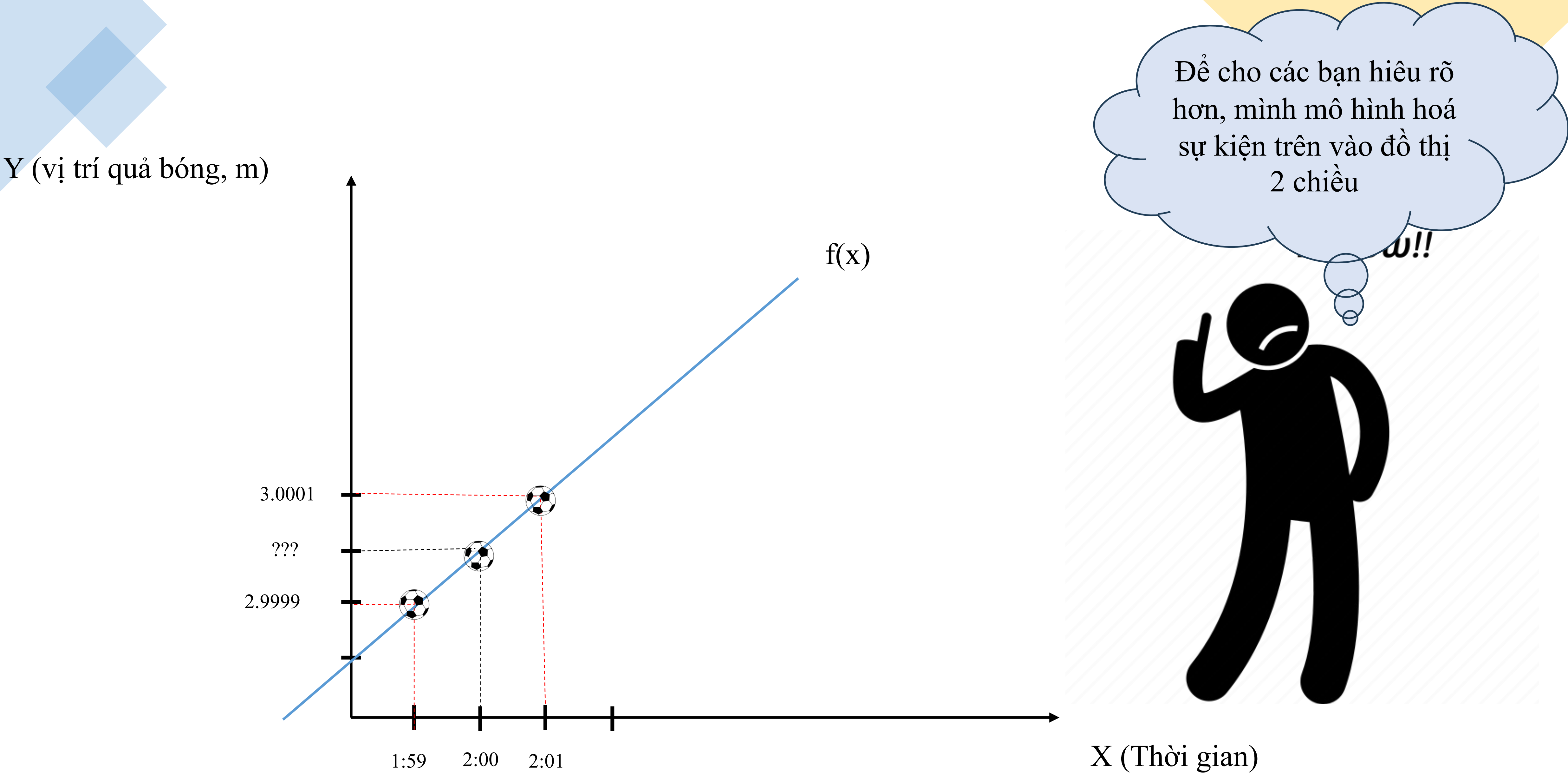
Điều gì đang diễn ra ở phút thứ 2:00 nhỉ?

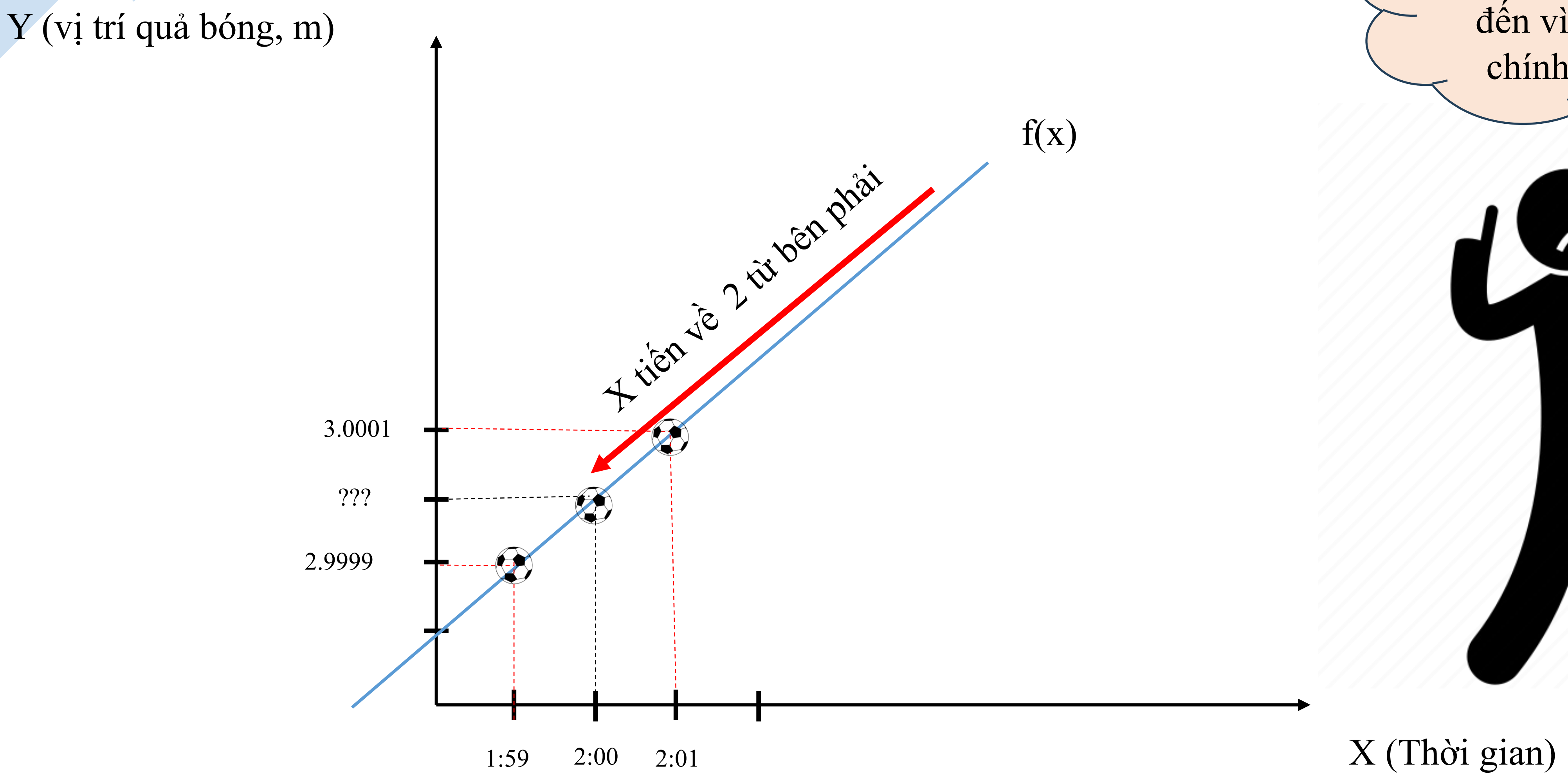


Phút 1:59.999 quả bóng ở vị trí 2.999m. Và phút 2:00.001 quả bóng ở vị trí 3.0001 mét.

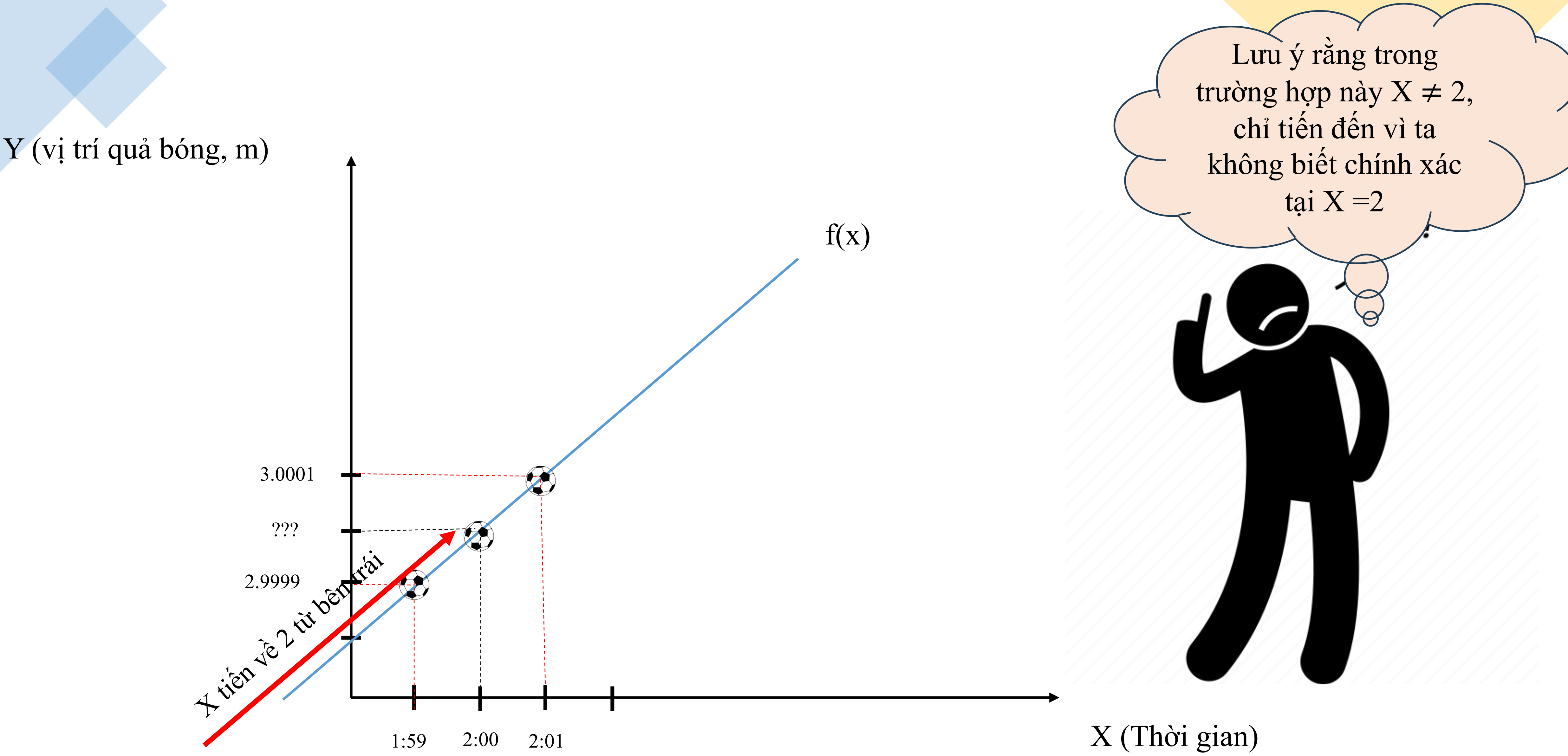
Như vậy, Phút 2, quả bóng có thể ở cự ly mét $\sim 3m$.

Giả sử bạn đang xem một trận bóng đá, nhưng kết nối không ổn định lắm, và kết quả như hình trên, các bạn bị mất khung hình ở thời gian 4:00 phút


Y (vị trí quả bóng, m)
f(x)
3.0001
???
2.9999
1:59
2:00
2:01
X (Thời gian)
Để cho các bạn hiêu rõ hơn, mình mô hình hoá sự kiện trên vào đồ thị 2 chiều


Y (vị trí quả bóng, m)
f(x)
X tiến về 2 từ bên phải
3.0001
???
2.9999
1:59
2:00
2:01
X (Thời gian)

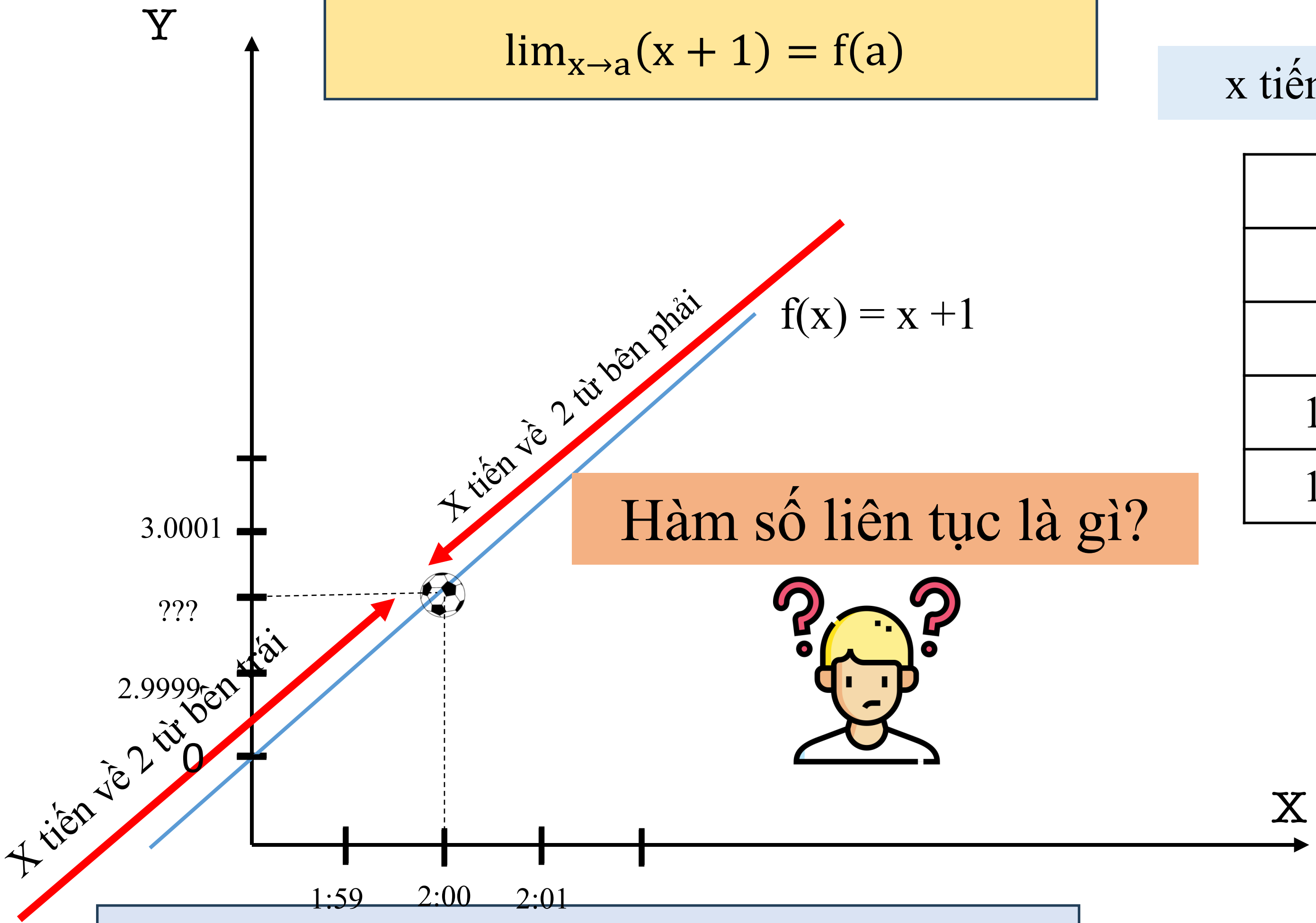

Lưu ý rằng trong trường hợp này X không bằng 2, chỉ tiến đến vì ta không biết chính xác tại X =2


Y (vị trí quả bóng, m)
f(x)
3.0001
???
2.9999
X tiến về 2 từ bên trái
1:59
2:00
2:01
X (Thời gian)
Lưu ý rằng trong trường hợp này X ≠ 2, chỉ tiến đến vì ta không biết chính xác tại X =2

Hàm số liên tục là gì?

Chỉ đúng với **hàm số liên tục**

$$\lim_{x\to a}(x+1) = f(a)$$



Khi x tiến đến 2, giá trị f(x) là gì?
x chỉ tiến đến 2, chứ không bằng 2.
Giả sử f(x) = x +1

x tiến về 2 từ bên trái

x	F(x)
0	1
1.3	2.3
1.77	2.77
1.99	2.99

x tiến về 2 từ bên bên phải

x	F(x)
3	4
2.7	3.7
2.4	3.4
2.01	3.01

Khi x tiến đến 2 => f(x) đều tiến đến 3

Giới hạn (limit) của hàm f(x) = x+1 là 3, khi x tiến đến 2

$$\lim_{x\to 2}(x+1) = f(2) = 3$$

Khi và chỉ khi

$$\lim_{x\to 2^+}(x+1) = 3 \text{ (phải)}$$
$$\lim_{x\to 2^-}(x+1) = 3 \text{ (trái)}$$

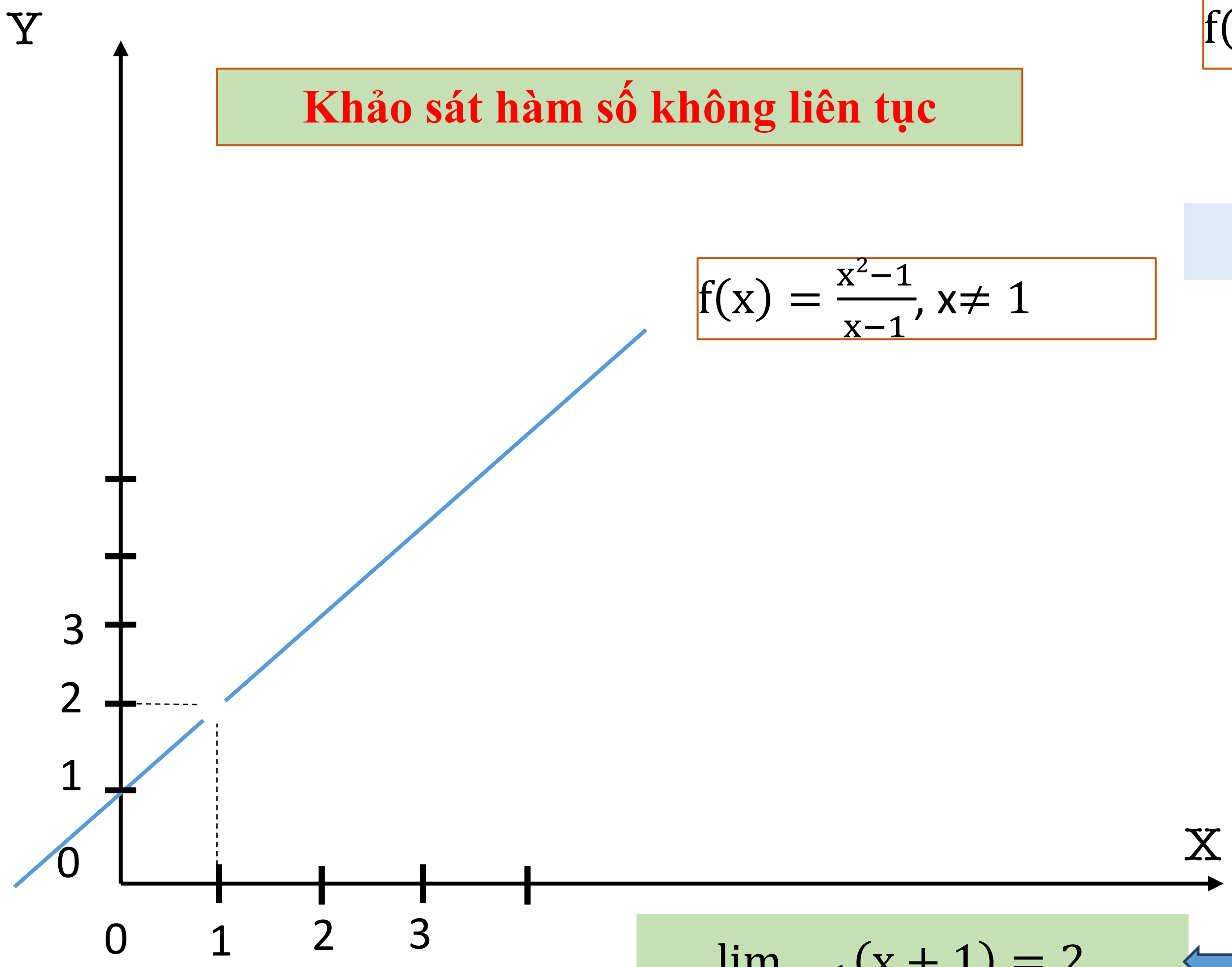
(Vô định)

$f(x) = \frac{x^2 - 1}{x - 1}$ → $f(x) = \frac{1 - 1}{1 - 1} = 0$ → $f(x) = \frac{x^2-1}{x-1}, x \neq 1$ → $\lim_{x\to 1} f(x) = ???$

$f(x) = \frac{x^2-1}{x-1} = x+1, x \neq 1$

Khảo sát hàm số không liên tục

$f(x) = \frac{x^2-1}{x-1}, x \neq 1$



x tiến đến 1 từ bên trái

x	F(x)
0.9	1.9
0.99	1.99
0.999	1.999

x tiến đến 2 từ bên bên phải

x	F(x)
1.1	2.1
1.01	2.01
1.001	2.001

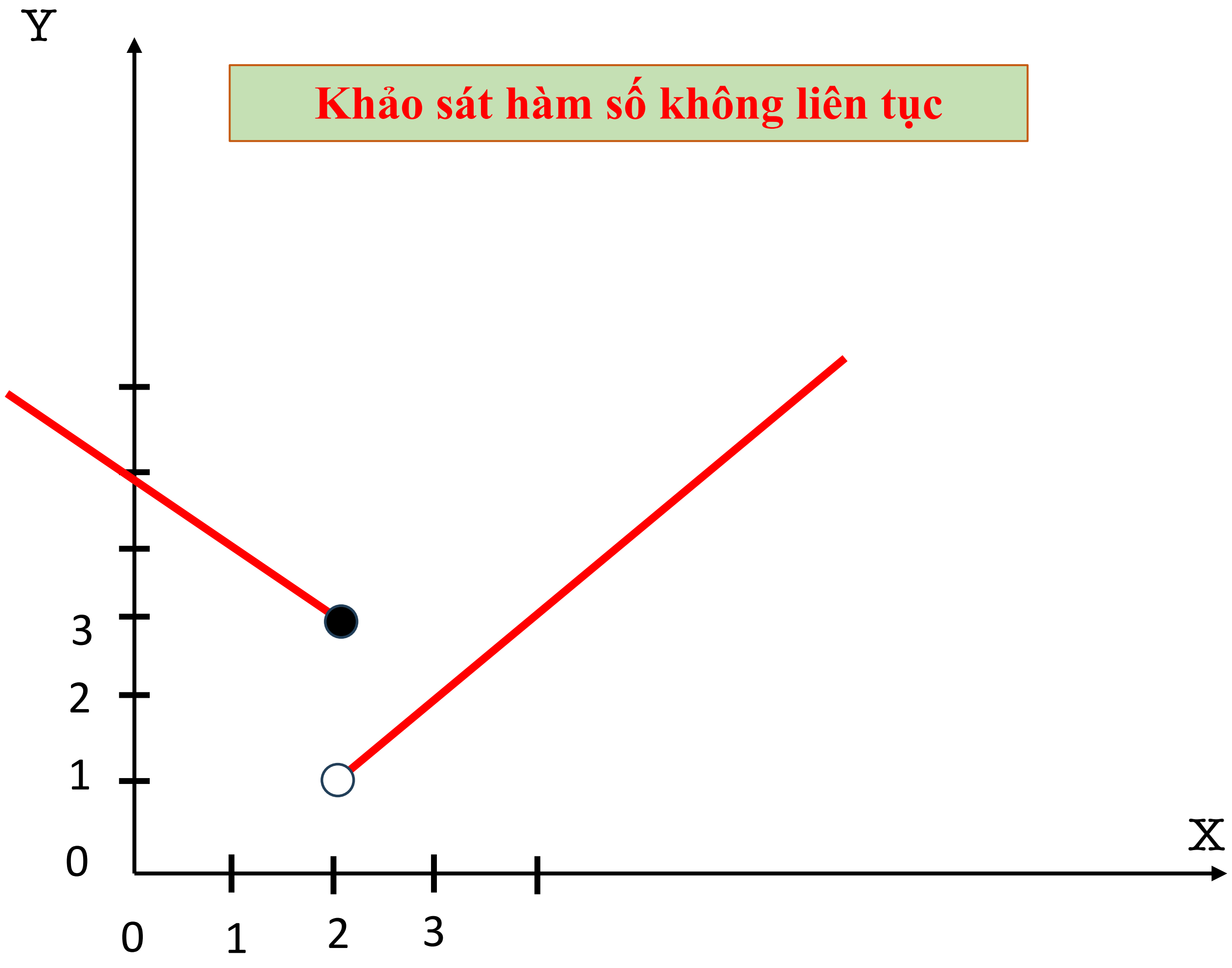
Khi x tiến đến 1 => F(x) đều tiến đến 2

$\lim_{x\to 1^+}(x + 1) = 2$ (phải)
$\lim_{x\to 1^-}(x + 1) = 2$ (trái)

$\lim_{x\to 1}(x + 1) = 2$

$$f(x) = \begin{cases} -x + 5, \text{if } x \le 2 \\ x - 1, \text{if } x > 2 \end{cases}$$

Khảo sát hàm số không liên tục



$\lim_{x \to 2}(f(x)) = 2$

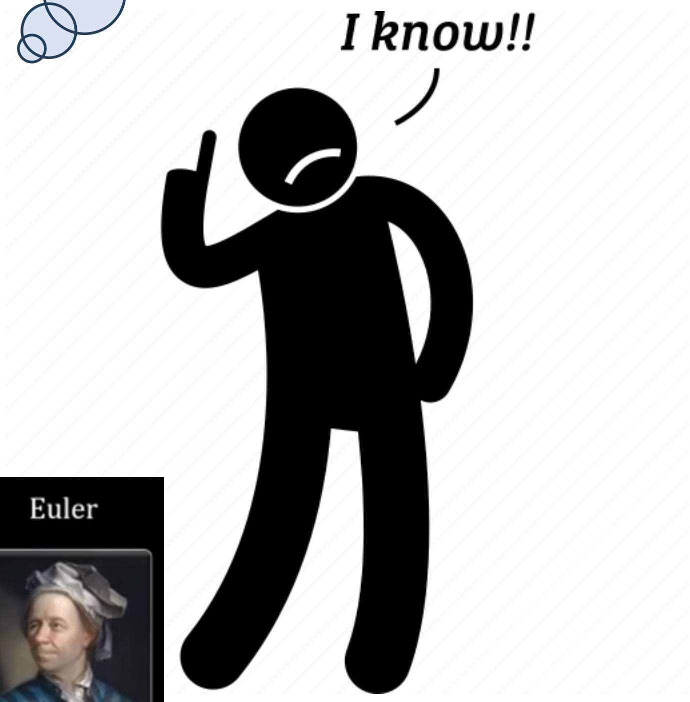
Không tồn tại

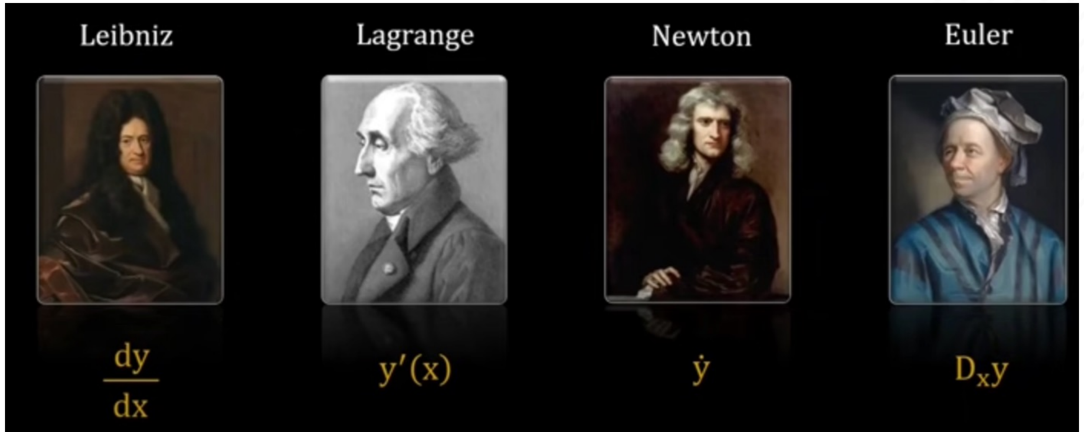
$\lim_{x \to 2-}(f(x)) = 3$ (phải)
$\lim_{x \to 2+}(x + 1) = 1$ (trái)

Đạo hàm rất quan trọng trong ML. Mình vẫn chưa hiểu đạo hàm là gì?

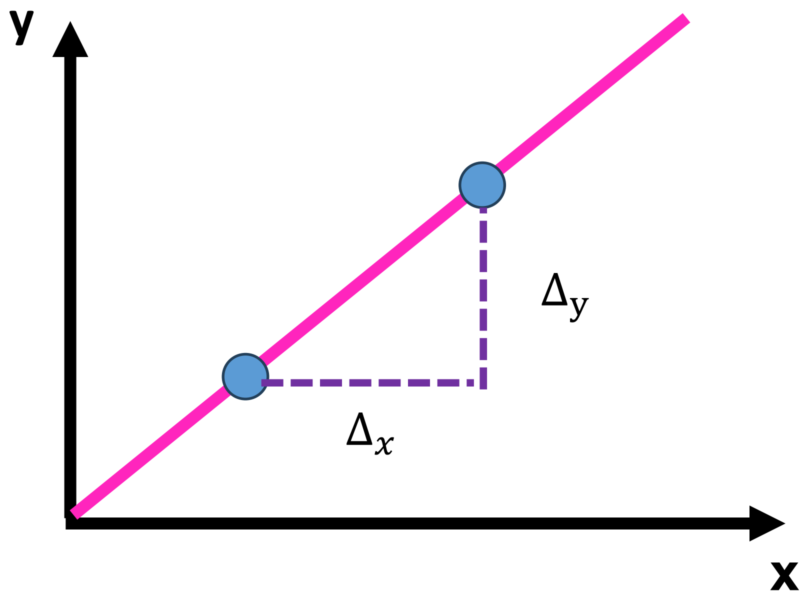
Để hiểu đạo hàm bần cần biết:

- Limits of function (giới hạn hàm số)
- Slope of the function (độ đốc hàm số)
- Secant line (đường cát tuyến)
- Targent line (đường tiếp tuyến)

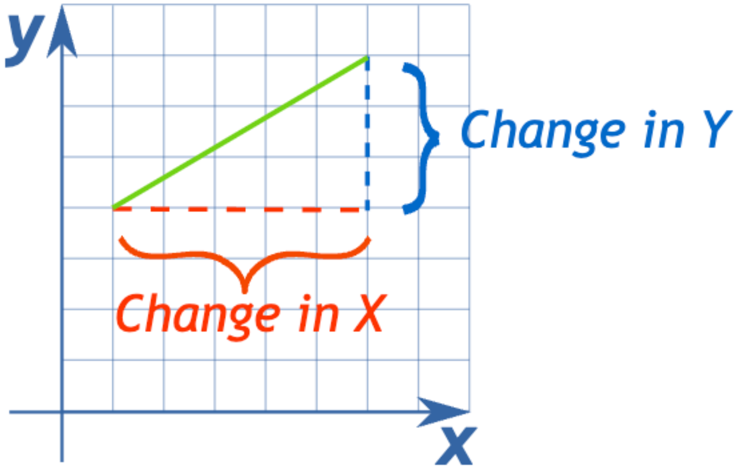




Slope of the Line



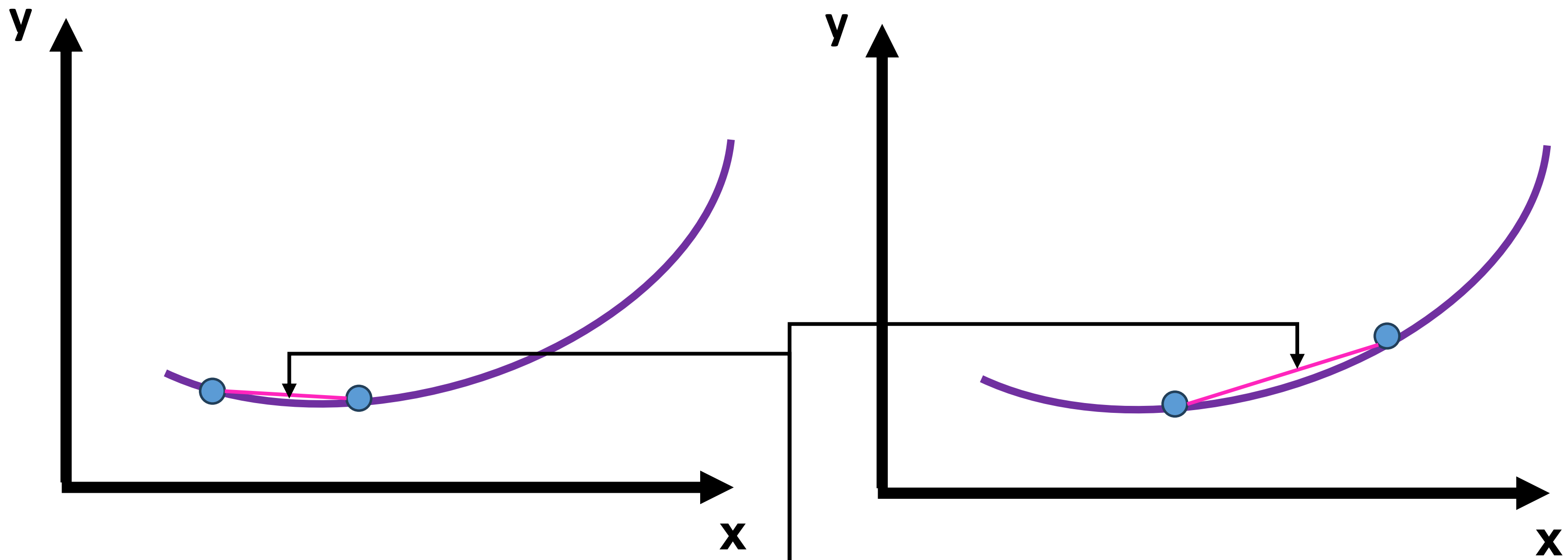
$$\text{Slope} = \frac{\text{Change in Y}}{\text{Change in X}}$$



Slope of the line = $\frac{\Delta_y}{\Delta_x}$ Độ dốc đường thẳng $\frac{\Delta_y}{\Delta_x}$

Nếu chúng ta chọn ngẫu nhiên 2 điểm bất kỳ trên đường thẳng, chúng sẽ có cùng giá trị slope $\frac{\Delta x}{\Delta y}$ (constant rate of change)

Slope of the Curve



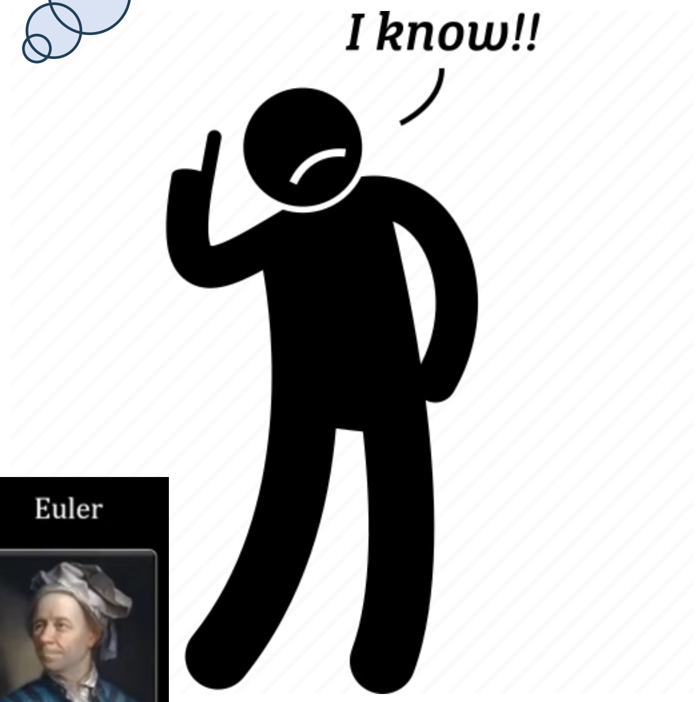
Slope of the Secant line (đường cát tuyến)

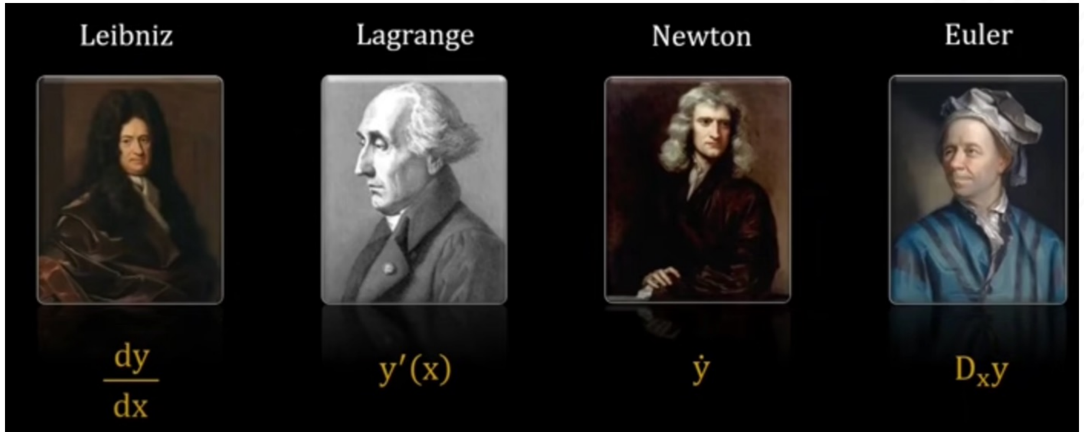
Khi lấy 2 điểm bất kỳ trên đường cong, Slope là khác nhau
Khi lấy 2 điểm bất kỳ trên đường thẳng, Slope là giống nhau

Đạo hàm rất quan trọng trong ML. Mình vẫn chưa hiểu đạo hàm là gì?

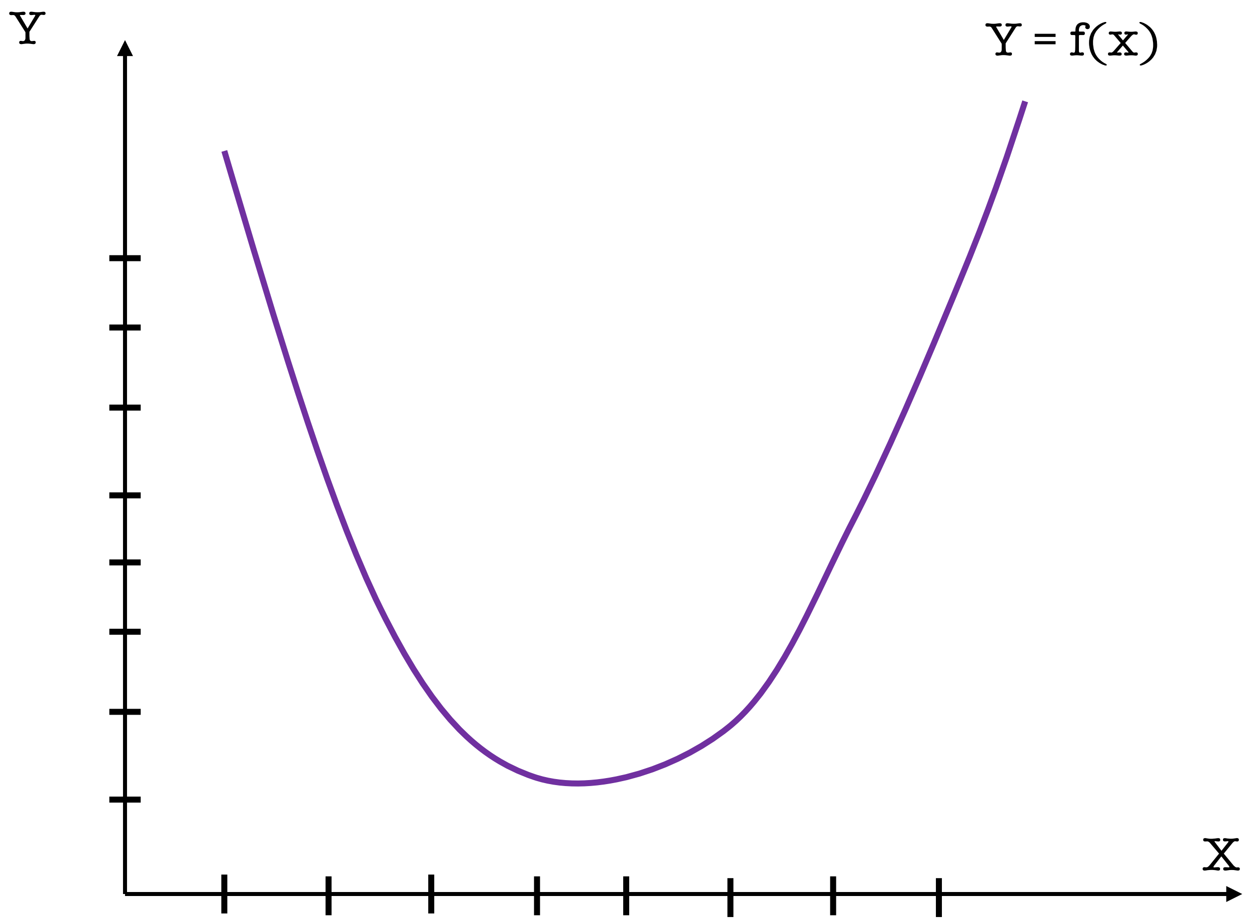
Để hiểu đạo hàm bần cần biết:

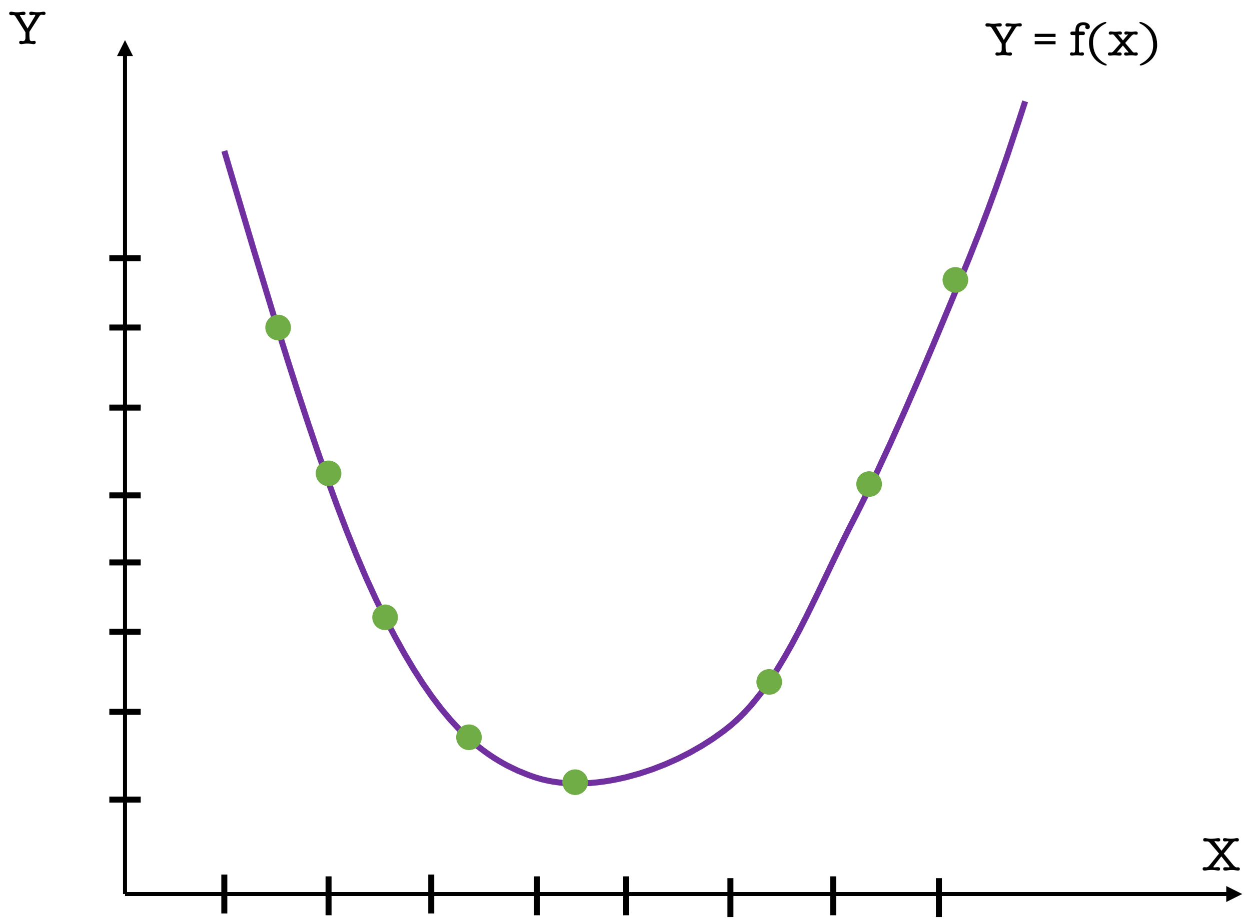
- Limits of function (giới hạn hàm số)
- Slope of the function (độ đốc hàm số)
- Secant line (đường cát tuyến)
- Targent line (đường tiếp tuyến)



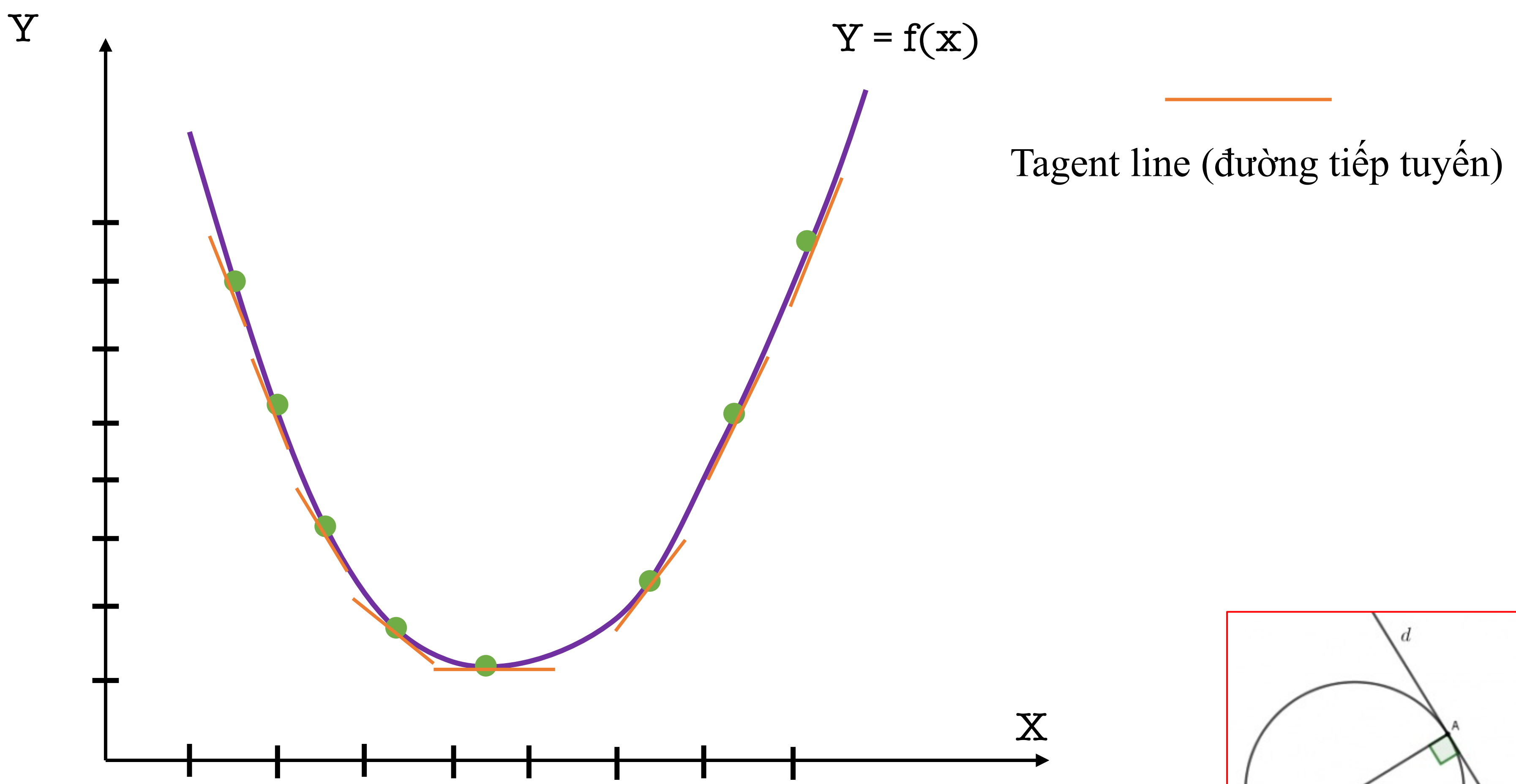


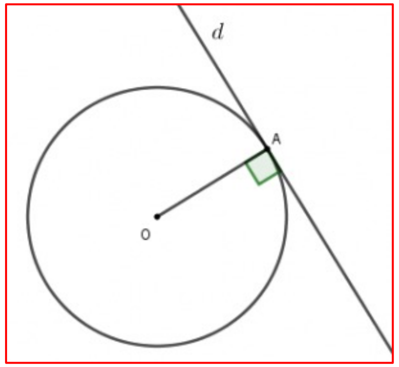
Tagent Line



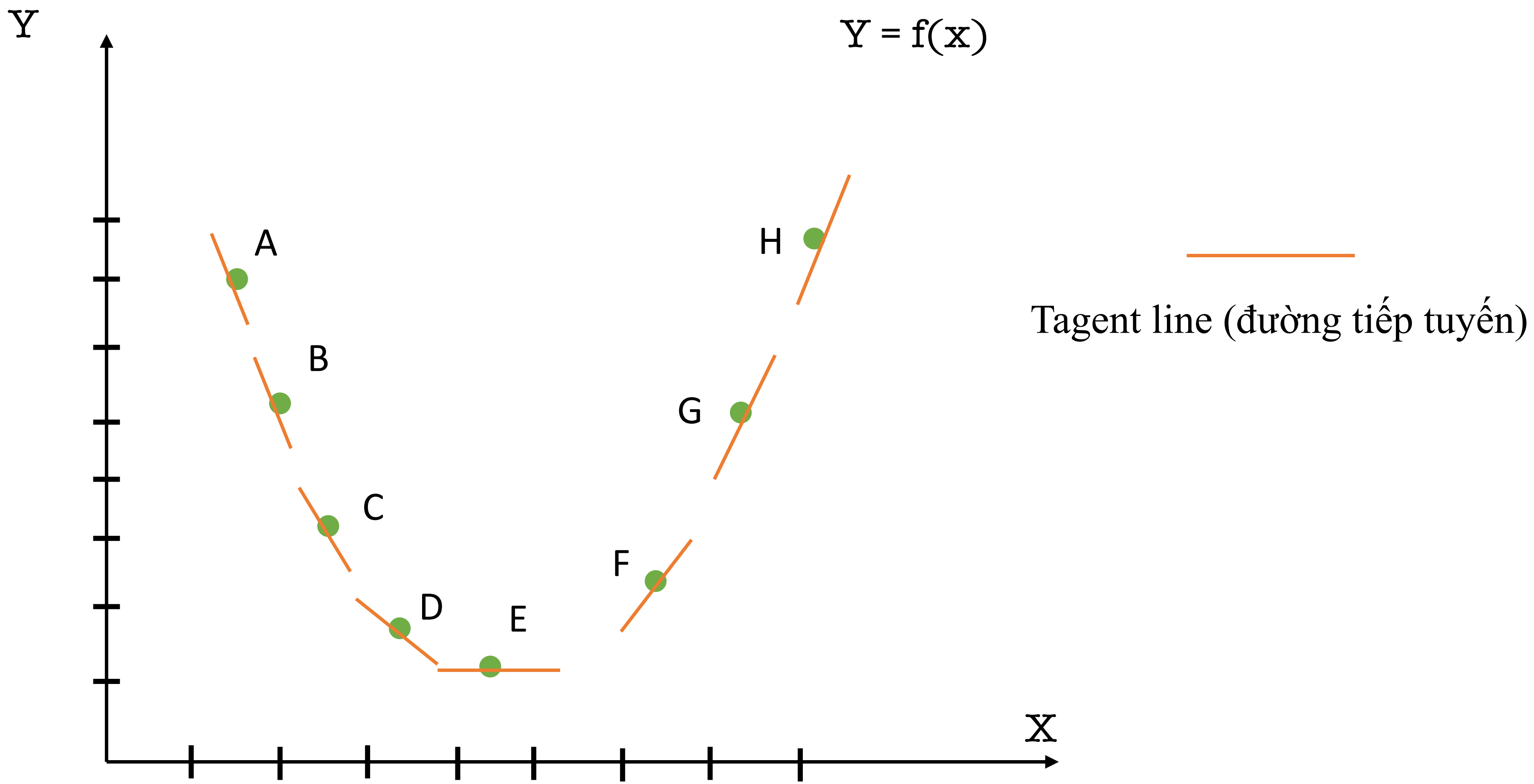


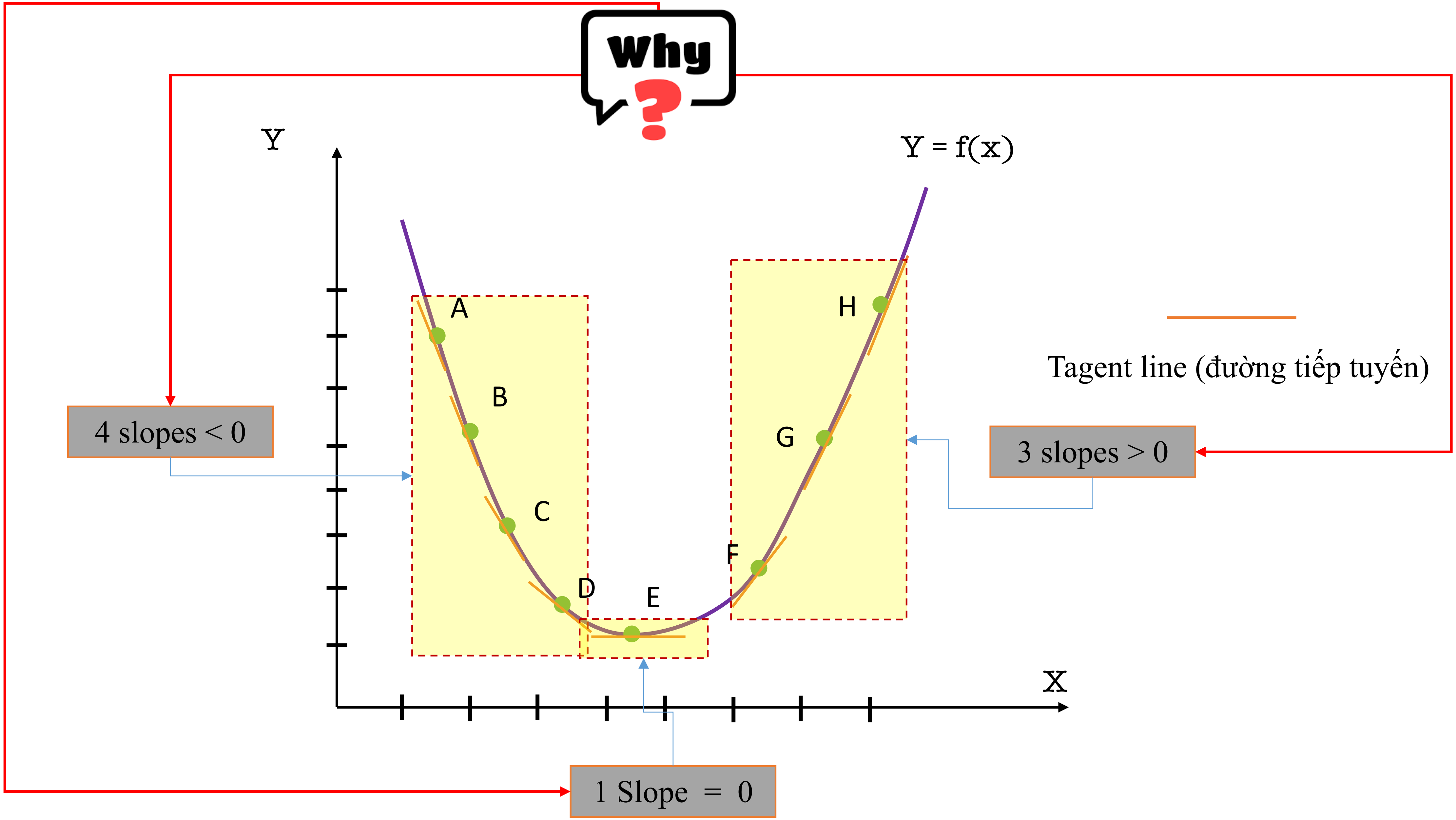
Y
Y = f(x)
X



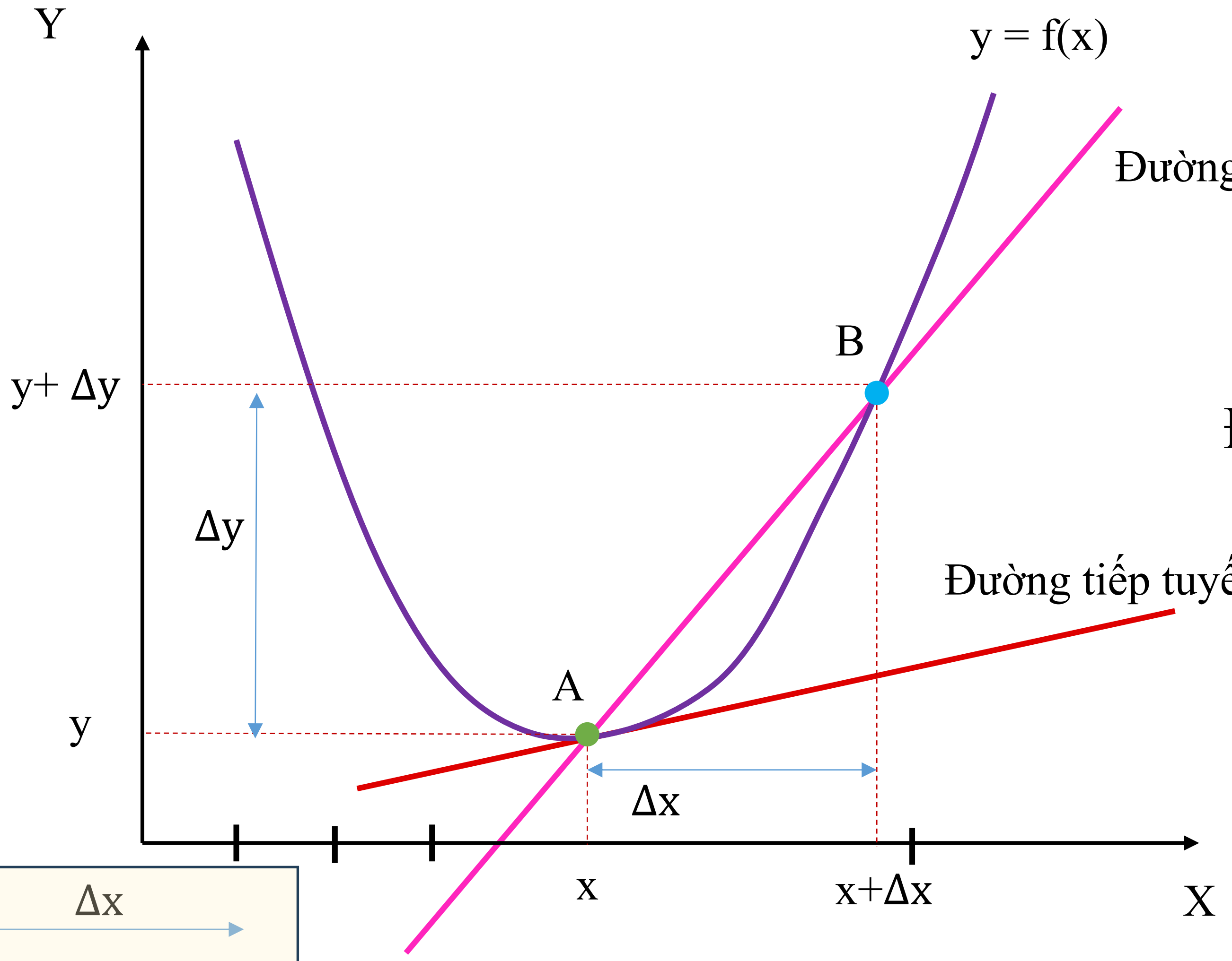


Tiếp tuyến (tangent) là đường thẳng chỉ tiếp xúc, chứ không cắt đồ thị tại một điểm nhất đinh


Y
Y = f(x)
A
B
C
D
E
F
G
H
X
Tagent line (đường tiếp tuyến)



Làm sao để tính độ dốc chính xác của các đường tuyếp tiến trên hàm y=f(x)



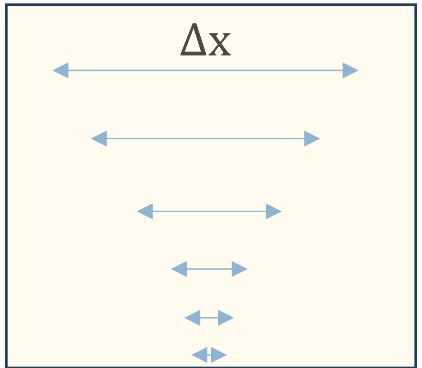
Độ dốc đường cát tuyến $= \dfrac{\Delta_y}{\Delta_x}$

$$\frac{\Delta_y}{\Delta_x} = \frac{f(x+\Delta_x)-f(x)}{\Delta_x}$$

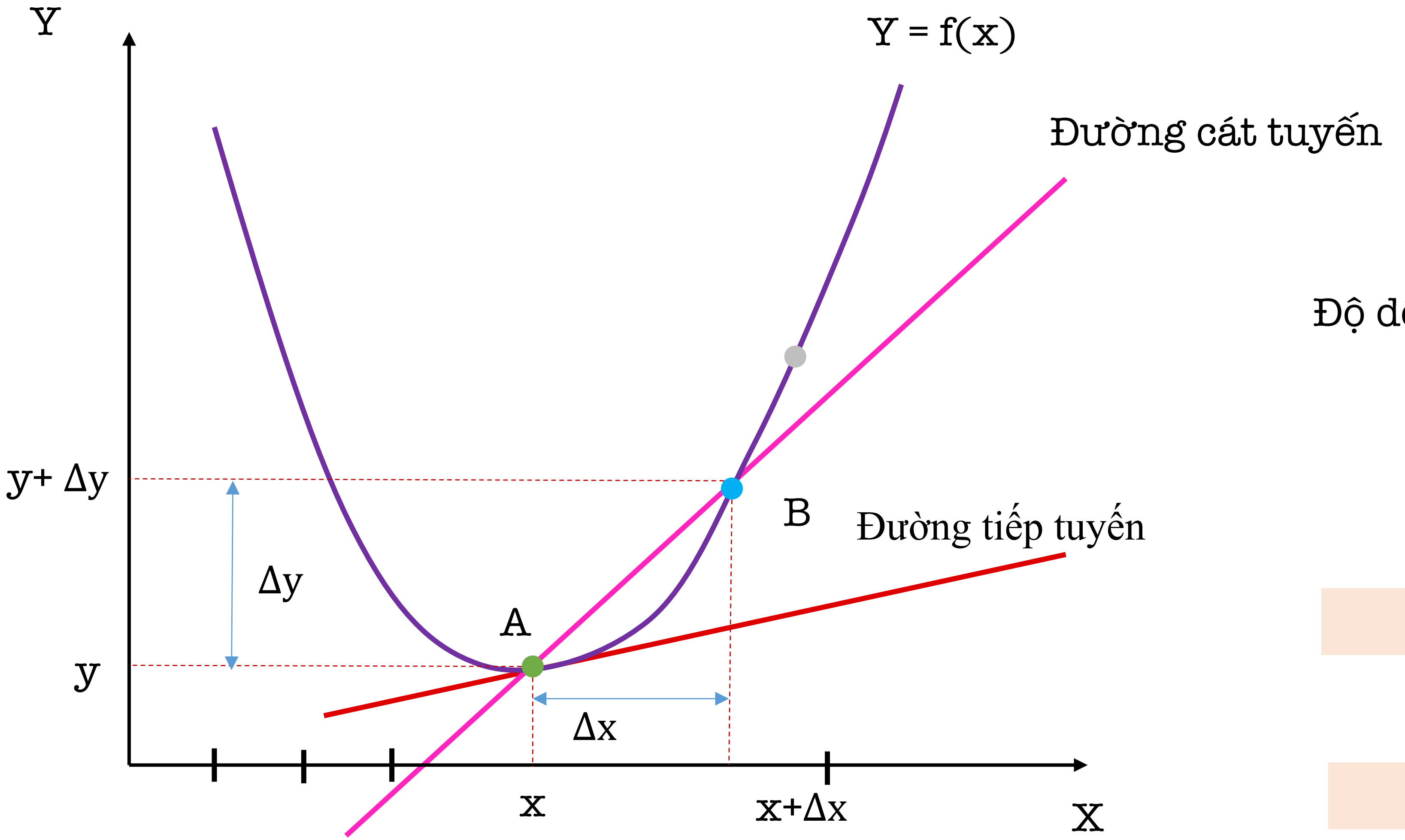
Giả sử rằng Δ_x tiến về 0

⇩

Hay nói các khác, Δ_x trở nên ngắn đi



Tại sao không giả sử $\Delta_x = 0$



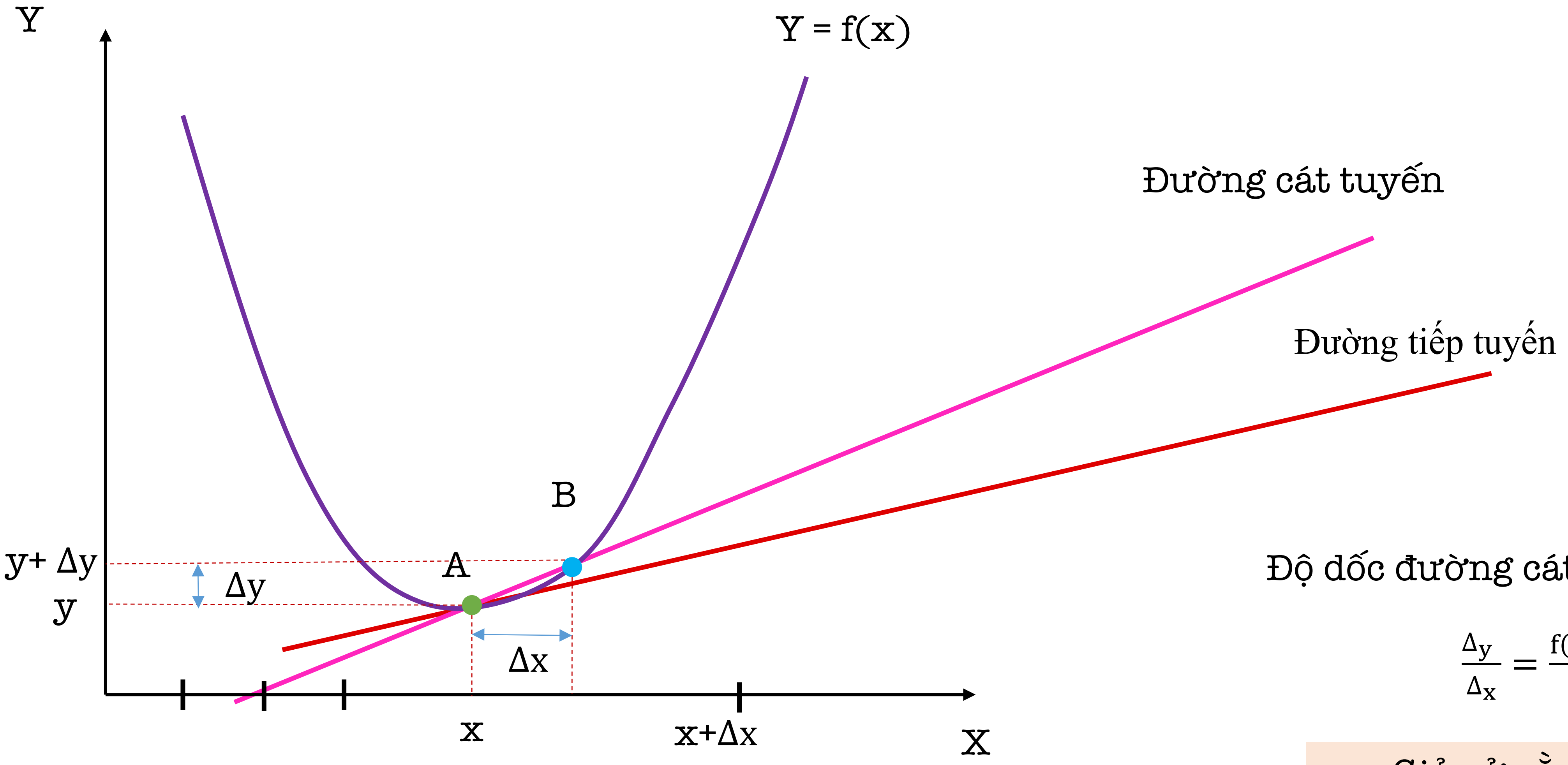
Độ dốc đường cát tuyến = $\frac{\Delta_y}{\Delta_x}$

$$\frac{\Delta_y}{\Delta_x} = \frac{f(x+\Delta_x)-f(x)}{\Delta_x}$$

Giả sử rằng Δ_x--> 0

Δ_x trở nên ngắn đi

Tại sao không giả sử $\Delta_x = 0$



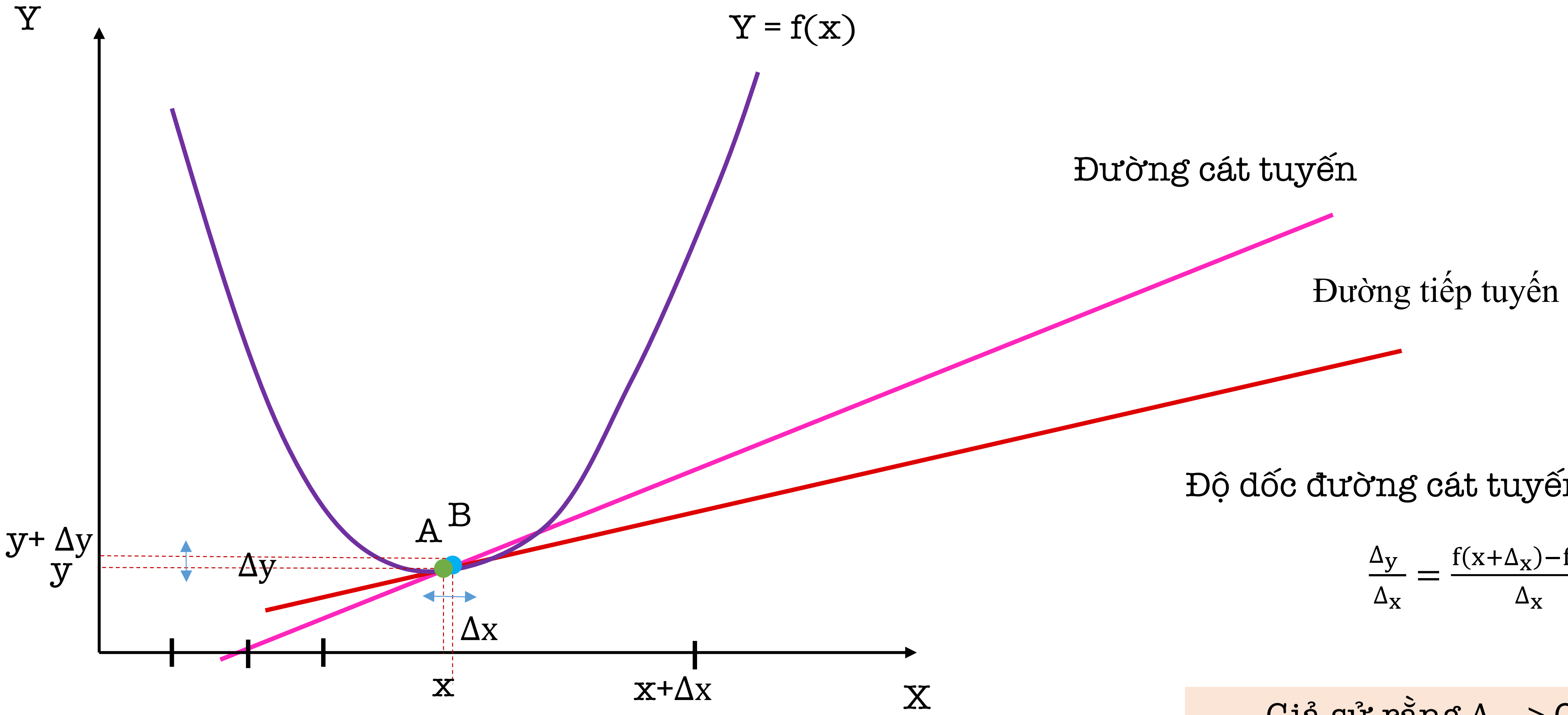
Độ dốc đường cát tuyến = $\frac{\Delta_y}{\Delta_x}$

$$\frac{\Delta_y}{\Delta_x} = \frac{f(x+\Delta_x)-f(x)}{\Delta_x}$$

Giả sử rằng Δ_x--> 0

Δ_x trở nên ngắn đi

Chuyện gì xảy ra nếu Δx cực kỳ nhỏ



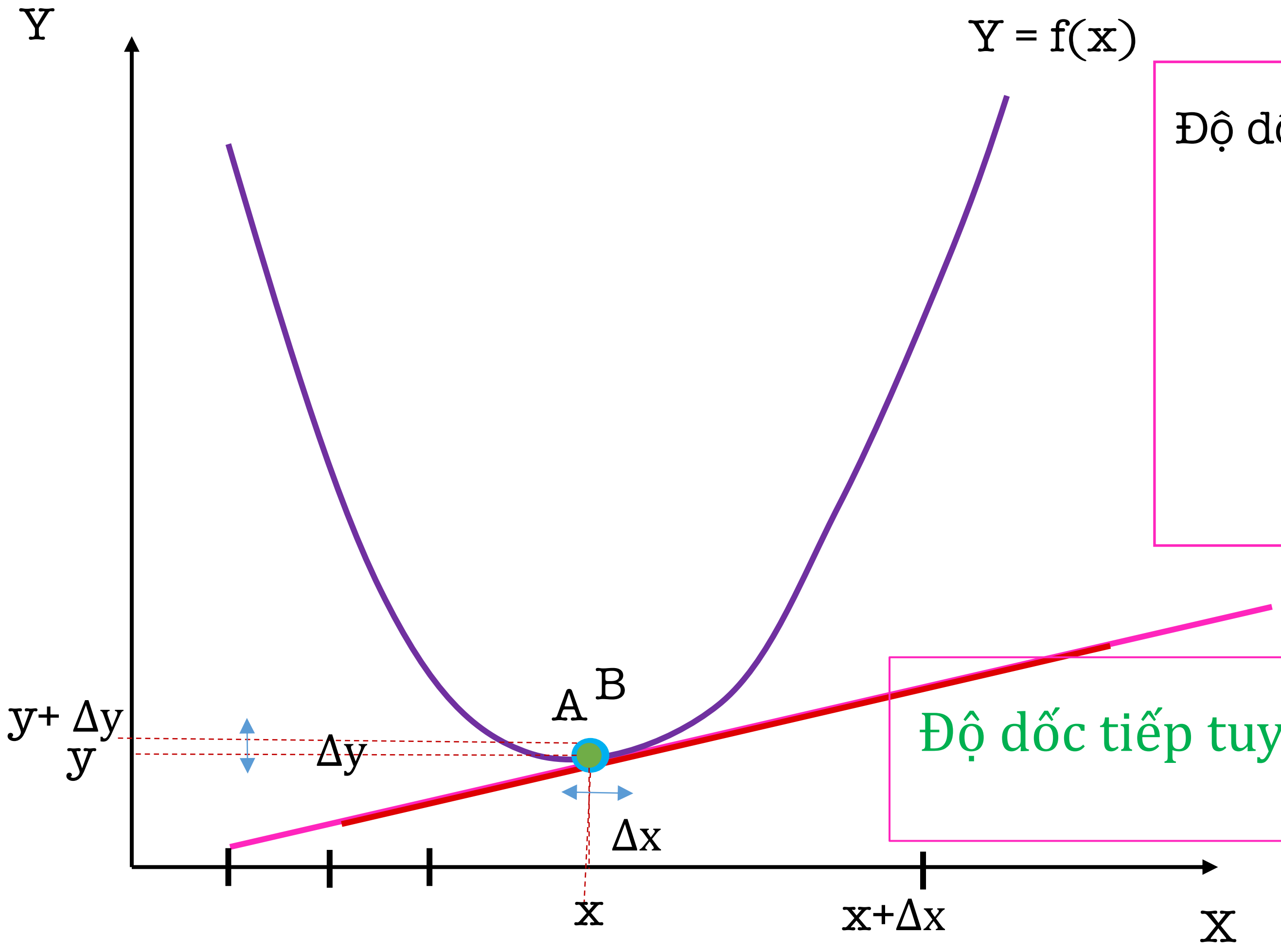
Độ dốc đường cát tuyến = $\frac{\Delta_y}{\Delta_x}$

$$\frac{\Delta_y}{\Delta_x} = \frac{f(x+\Delta_x)-f(x)}{\Delta_x}$$

Chuyện gì xảy ra nếu Δx cực kỳ nhỏ, và tiến về 0

Giả sử rằng Δ_x--> 0

Δ_x trở nên ngắn đi



Độ dốc đường cát tuyến = slope tiếp tuyến = $\frac{\Delta_y}{\Delta_x}$

$$\frac{\Delta_y}{\Delta_x} = \frac{f(x+\Delta_x)-f(x)}{\Delta_x}$$

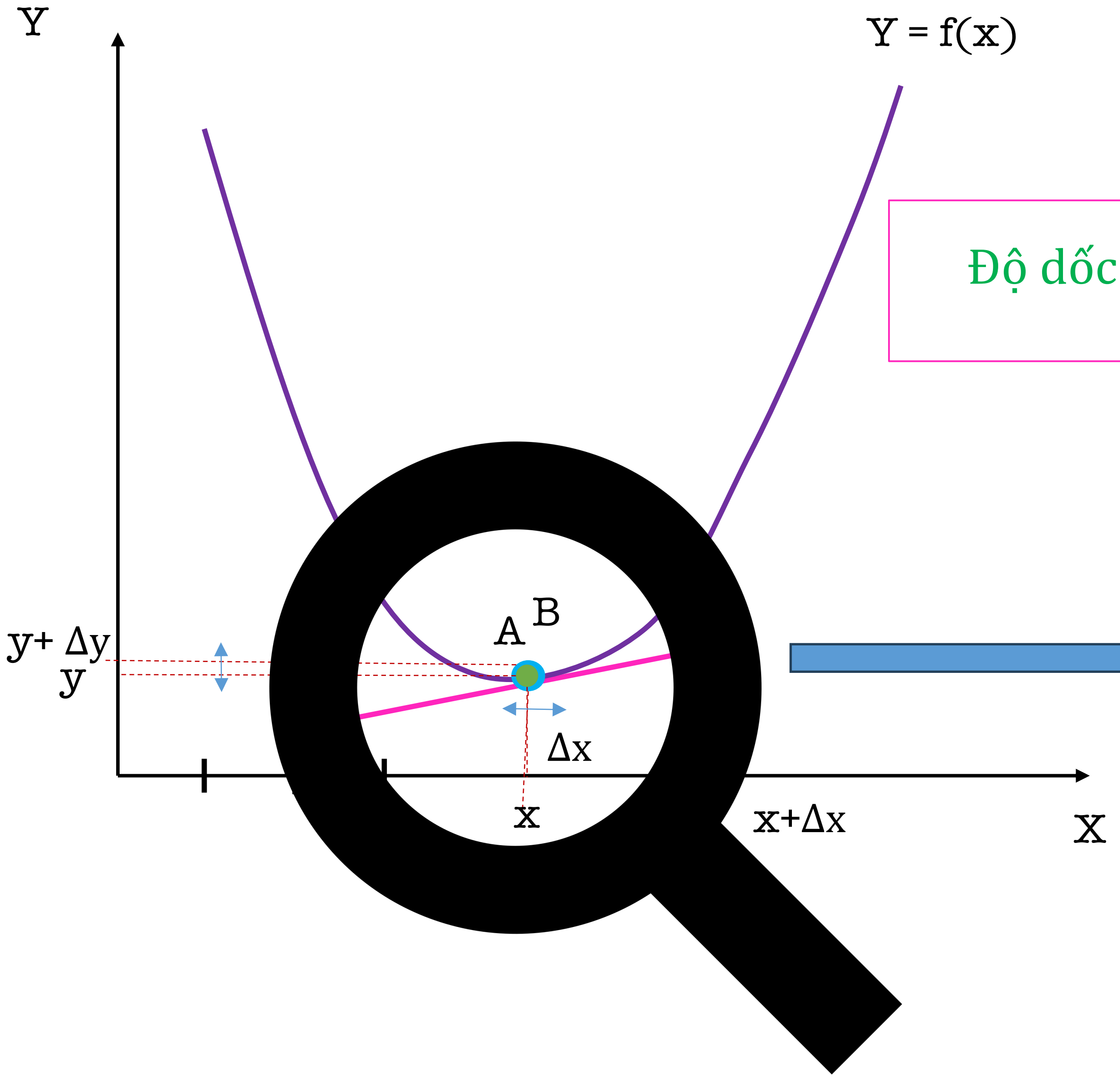
Giả sử rằng Δ_x tiến về 0

$$\text{Độ dốc tiếp tuyến} = \lim_{\Delta x \to 0} \frac{\Delta_y}{\Delta_x} = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

Đạo hàm/Derivative

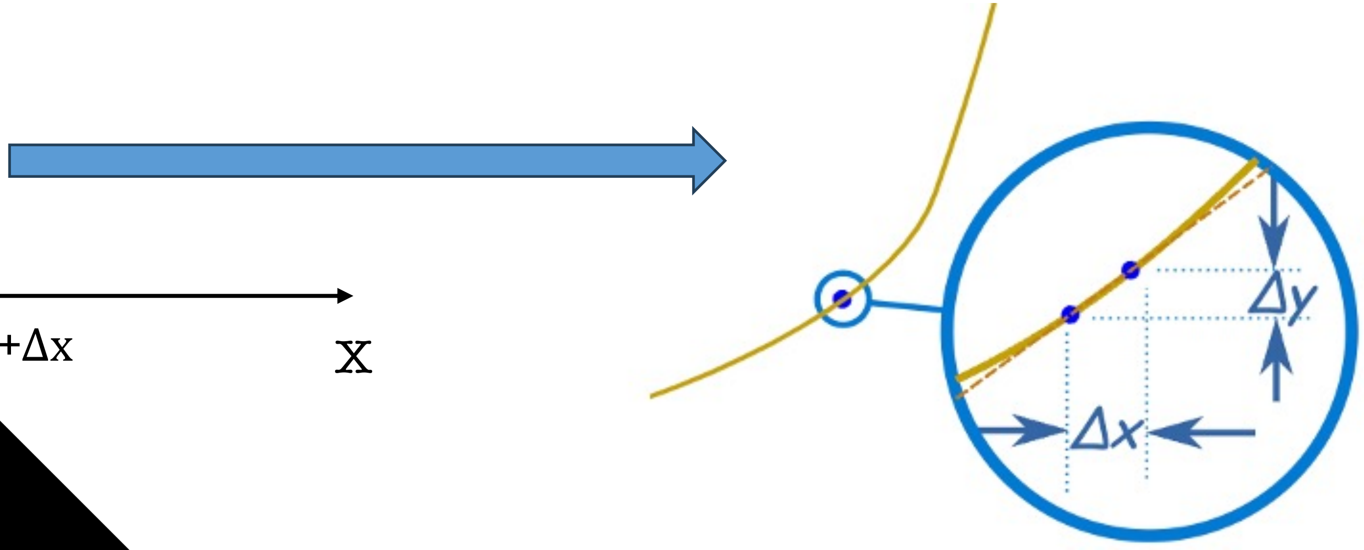
Chuyện gì xảy ra nếu Δx cực kỳ nhỏ, và tiến về 0

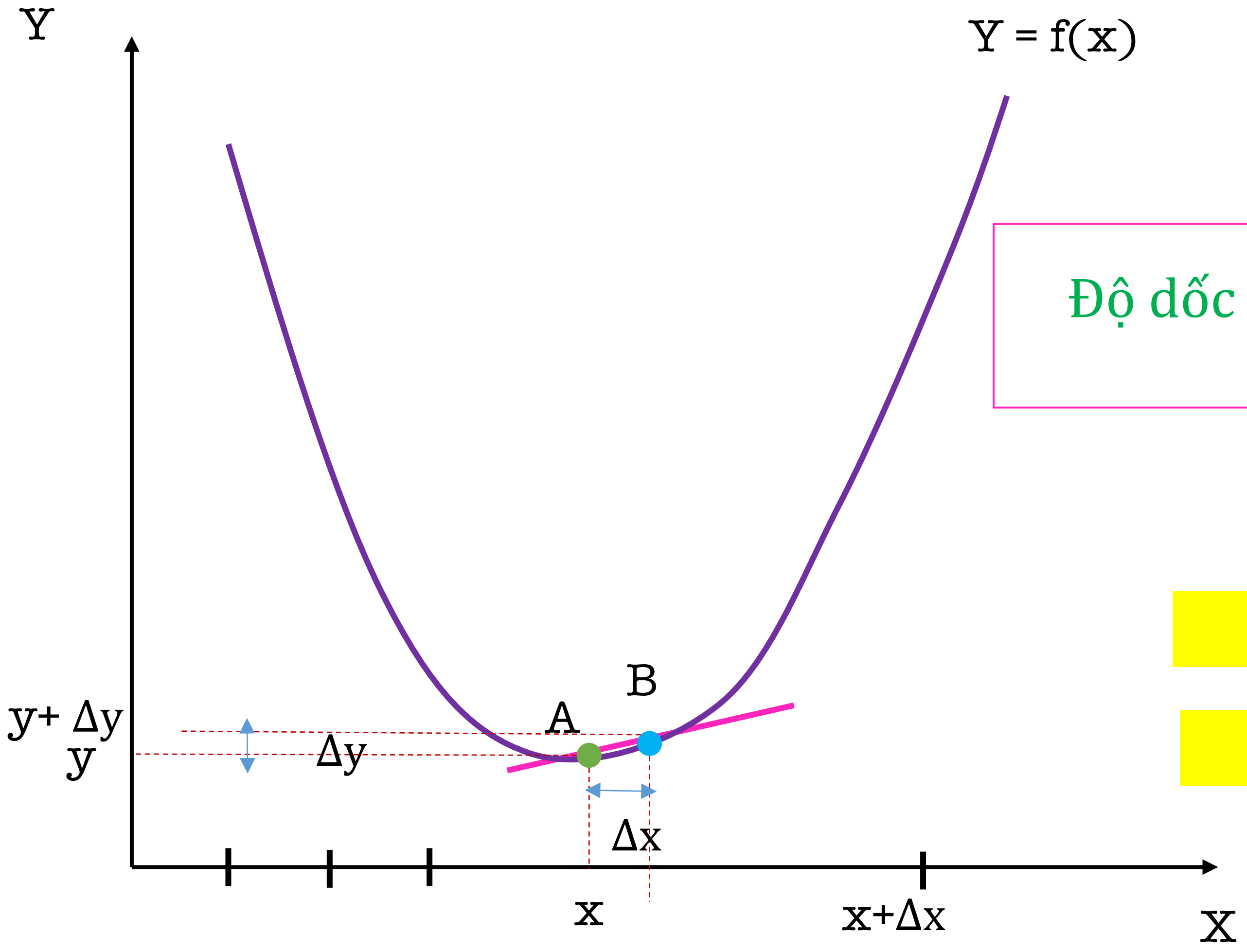
Đường cát tuyến sẽ trùng với đường tiếp tuyến



d: Differential/Vi phân

$$\text{Độ dốc cát tuyến} = \frac{\Delta x}{\Delta y} = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$$





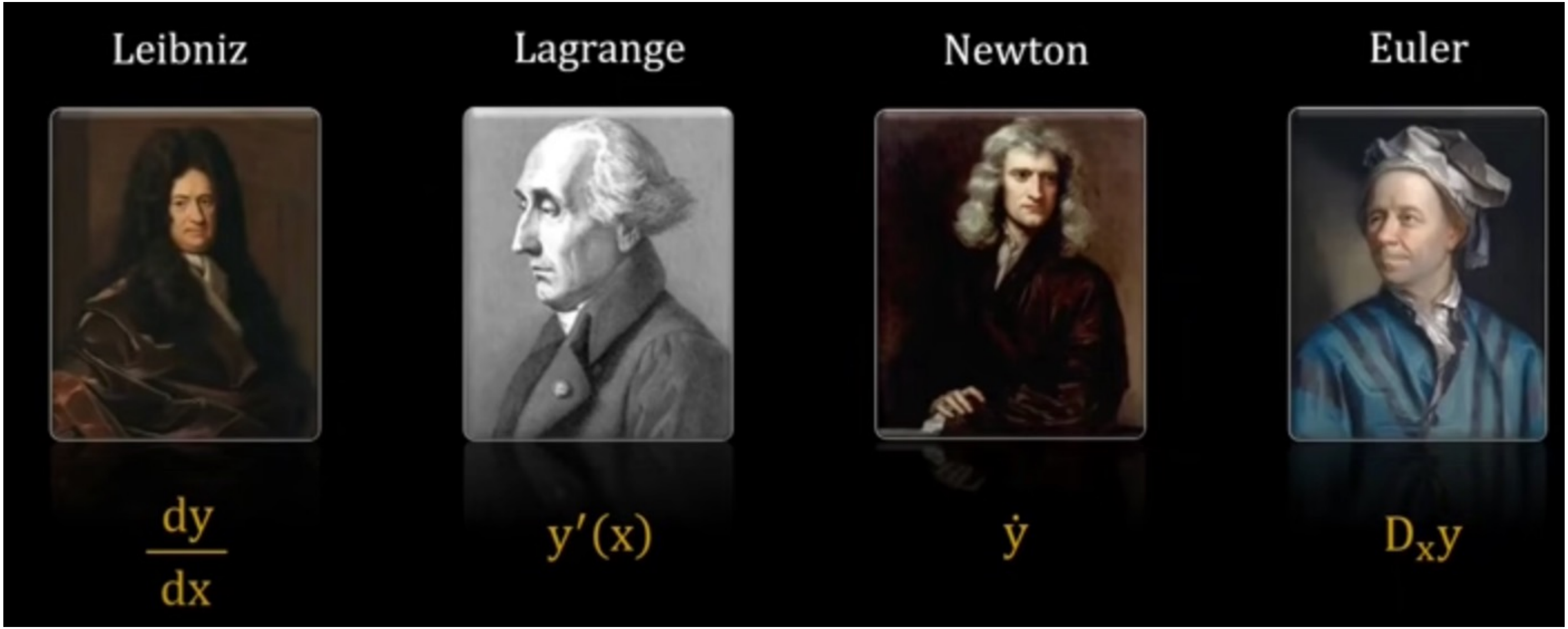
$$\text{Độ dốc tiếp tuyến} = \frac{d_y}{d_x} = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

Vì Δx siêu nhỏ nên ký hiệu là dx

Vì Δy siêu nhỏ nên ký hiệu là dy

Có bao nhiêu cách để ký hiệu đạo hàm?

How many ways to describe a derivative in Math?



Đức | Ý | Anh | Thuỵ Sĩ

Derivative

❖ Implementation

Cho hàm số f(x)

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

Công thức đạo hàm

$$f'(x) = 2x + 2$$

Công thức đạo hàm một bên

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

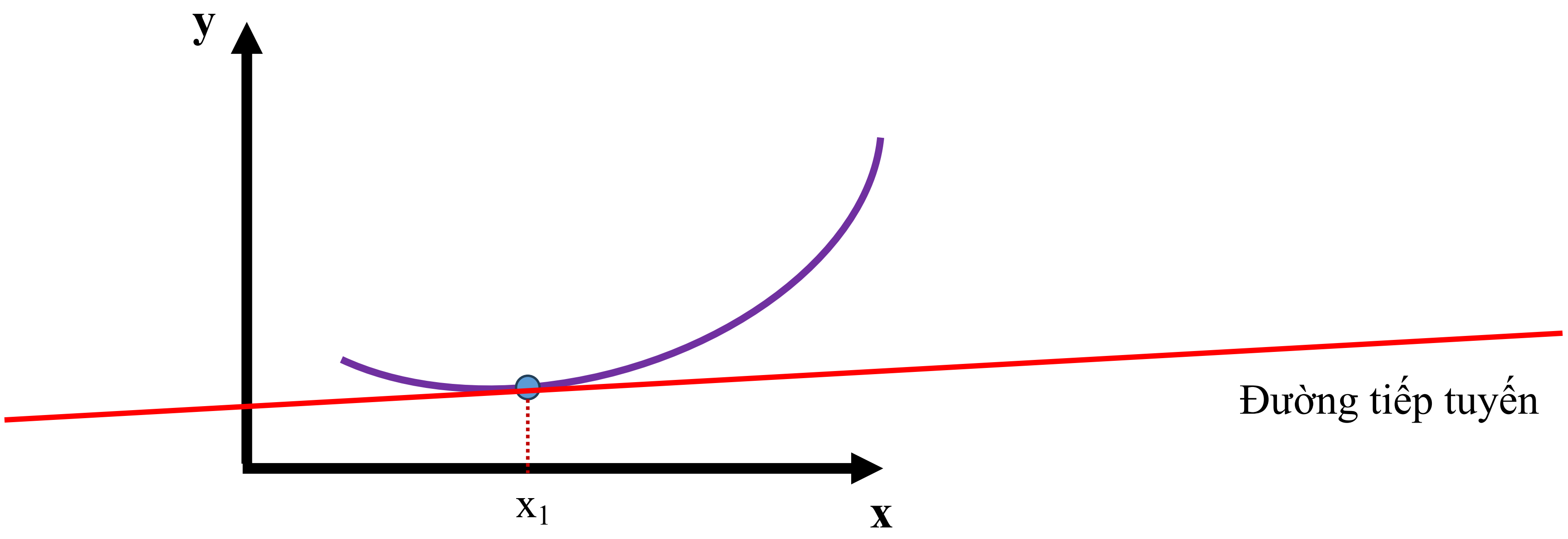
Công thức đạo hàm trung tâm

$$f'(x) = \lim_{\Delta x \to 0} \frac{f\left(x + \frac{\Delta x}{2}\right) - f\left(x - \frac{\Delta x}{2}\right)}{\Delta x}$$

```python
# đạo hàm trung tâm
def gradient(f, x, epsilon):
    return (f(x + epsilon/2) - f(x - epsilon/2)) / epsilon

def func(x):
    return x**2 + 2*x

print(f'(e=1.0e2 and x=2)  : {gradient(f=func, x=2.0, epsilon=1.0e2)}')
print(f'(e=1.0e1 and x=2)  : {gradient(f=func, x=2.0, epsilon=1.0e1)}')
print(f'(e=1.0e0 and x=2)  : {gradient(f=func, x=2.0, epsilon=1.0e0)}')
print(f'(e=1.0e-1 and x=2) : {gradient(f=func, x=2.0, epsilon=1.0e-1)}')
print(f'(e=1.0e-2 and x=2) : {gradient(f=func, x=2.0, epsilon=1.0e-2)}')
print(f'(e=1.0e-3 and x=2) : {gradient(f=func, x=2.0, epsilon=1.0e-3)}')
print(f'(e=1.0e-4 and x=2) : {gradient(f=func, x=2.0, epsilon=1.0e-4)}')
```

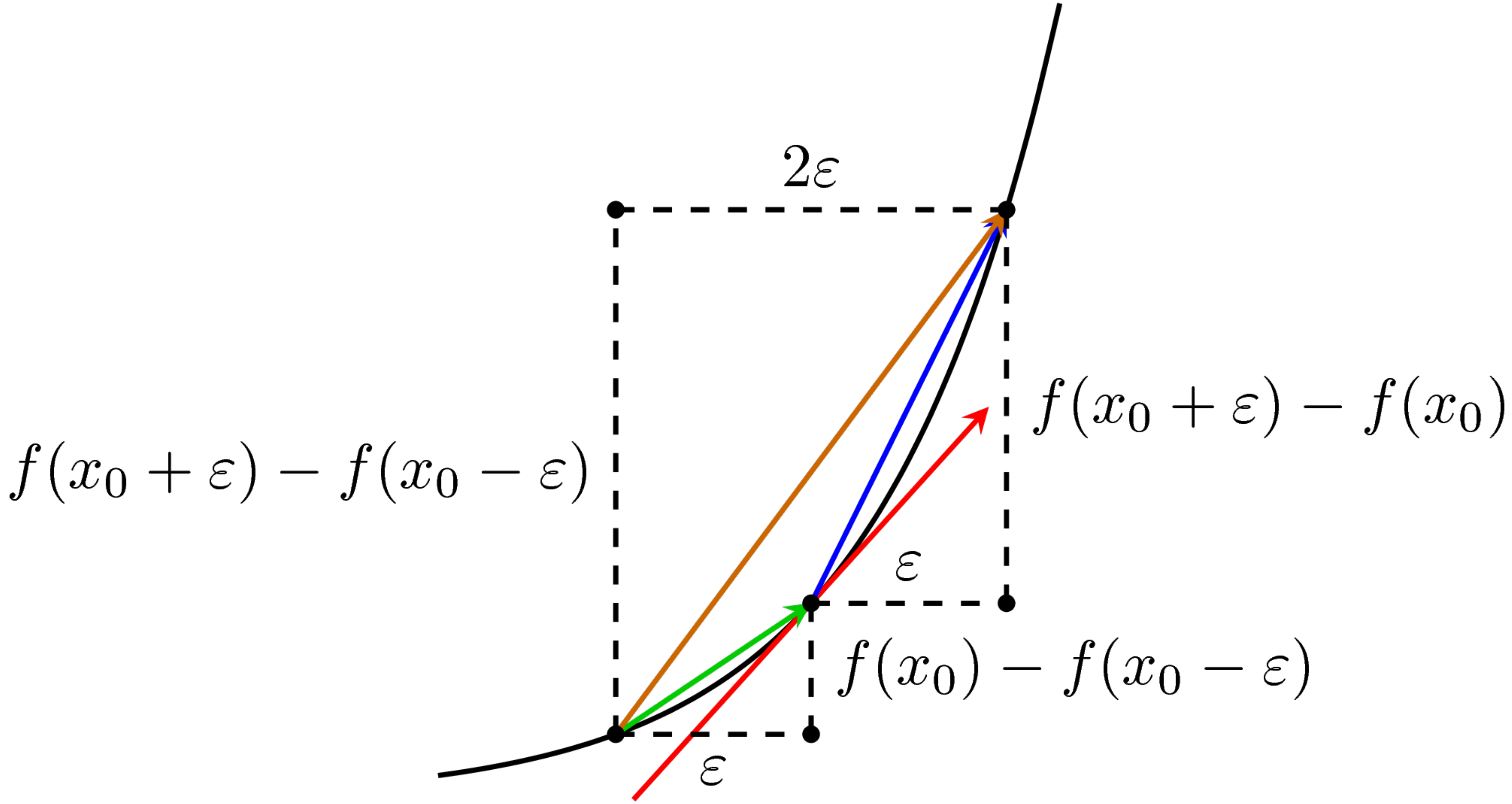
```
(e=1.0e2 and x=2)  : 6.0
(e=1.0e1 and x=2)  : 6.0
(e=1.0e0 and x=2)  : 6.0
(e=1.0e-1 and x=2) : 5.999999999999988
(e=1.0e-2 and x=2) : 5.999999999999783
(e=1.0e-3 and x=2) : 6.000000000011156
(e=1.0e-4 and x=2) : 6.00000000000378
```

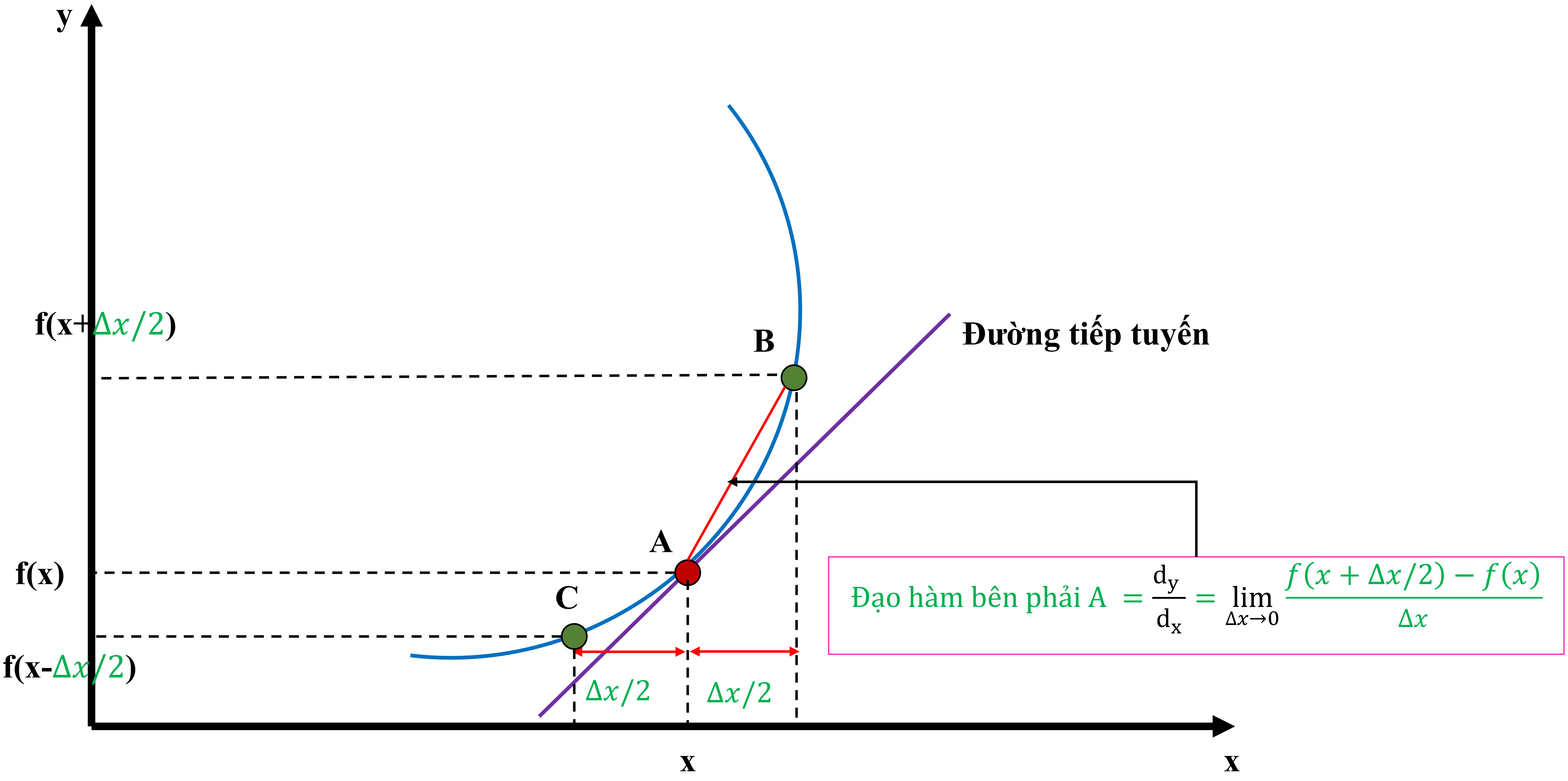
Đạo hàm trung tâm và **đạo hàm một bên**, cái nào phổ biến hơn?



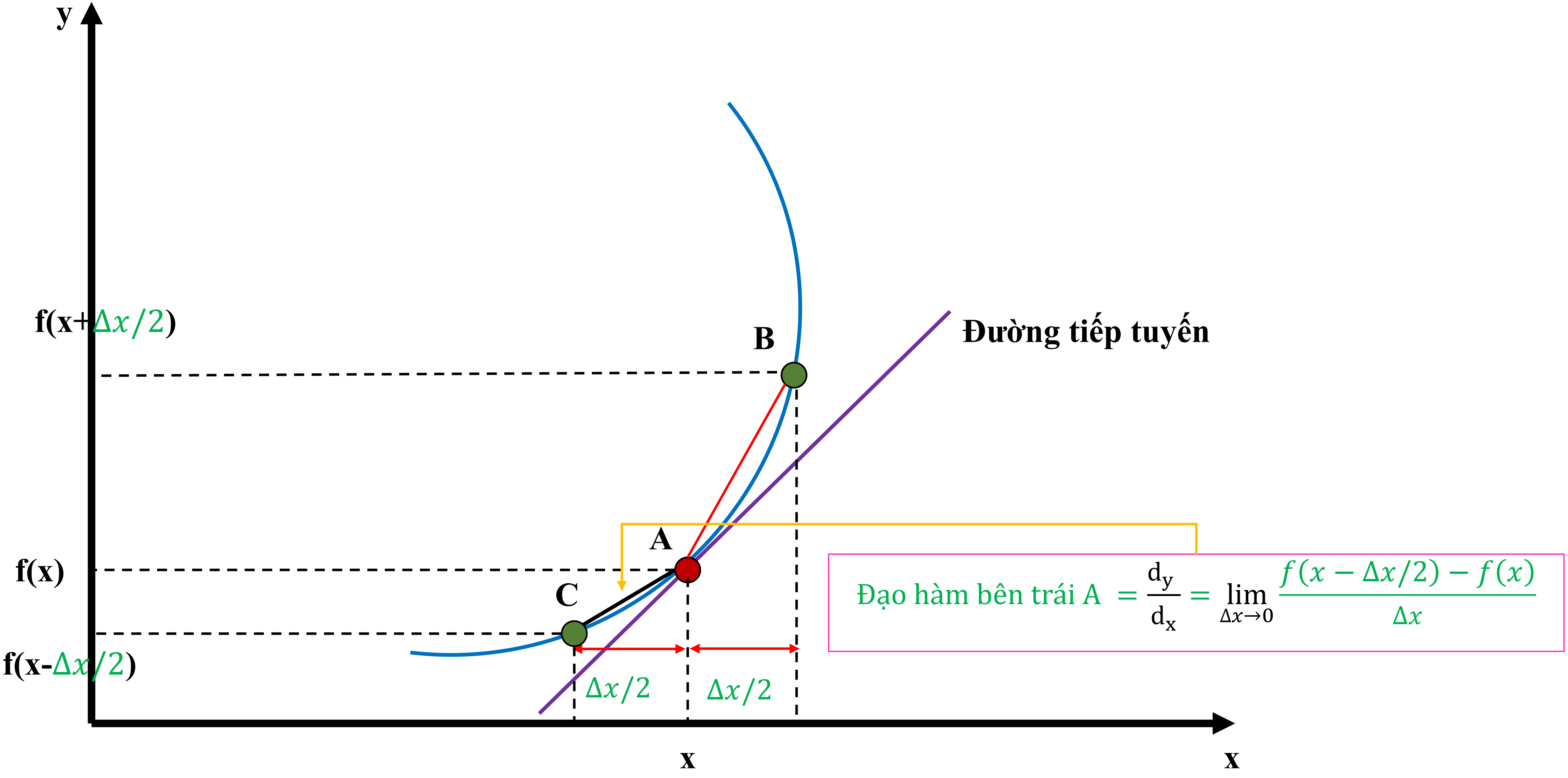
Đạo hàm tại điểm x1 chính là độ dốc (slope) của đường thẳng tiếp tuyến đi qua điểm x1

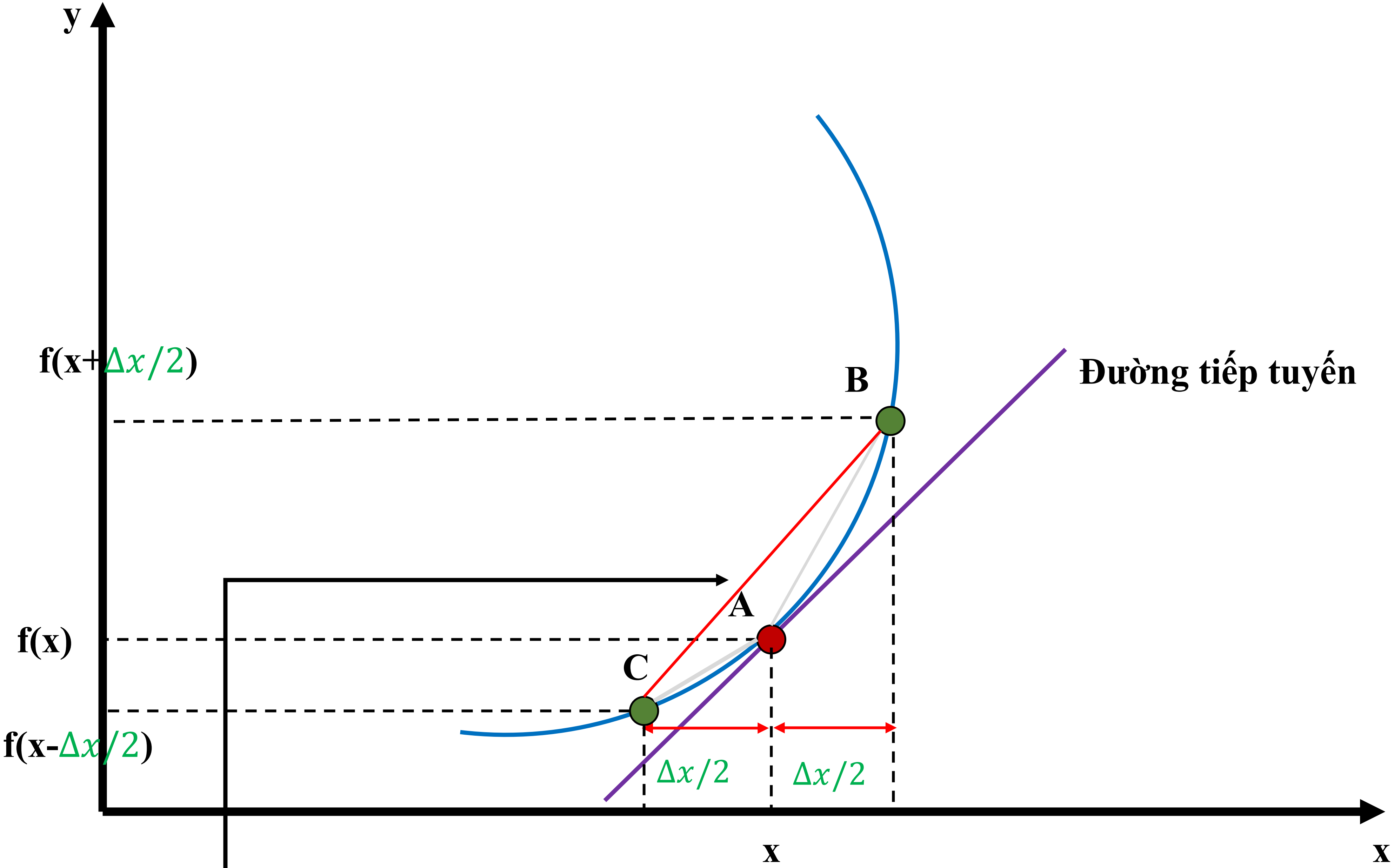
So sánh đạo hàm trung tâm và đạo hàm một bên, đạo hàm nào được dùng phổ biến hơn?



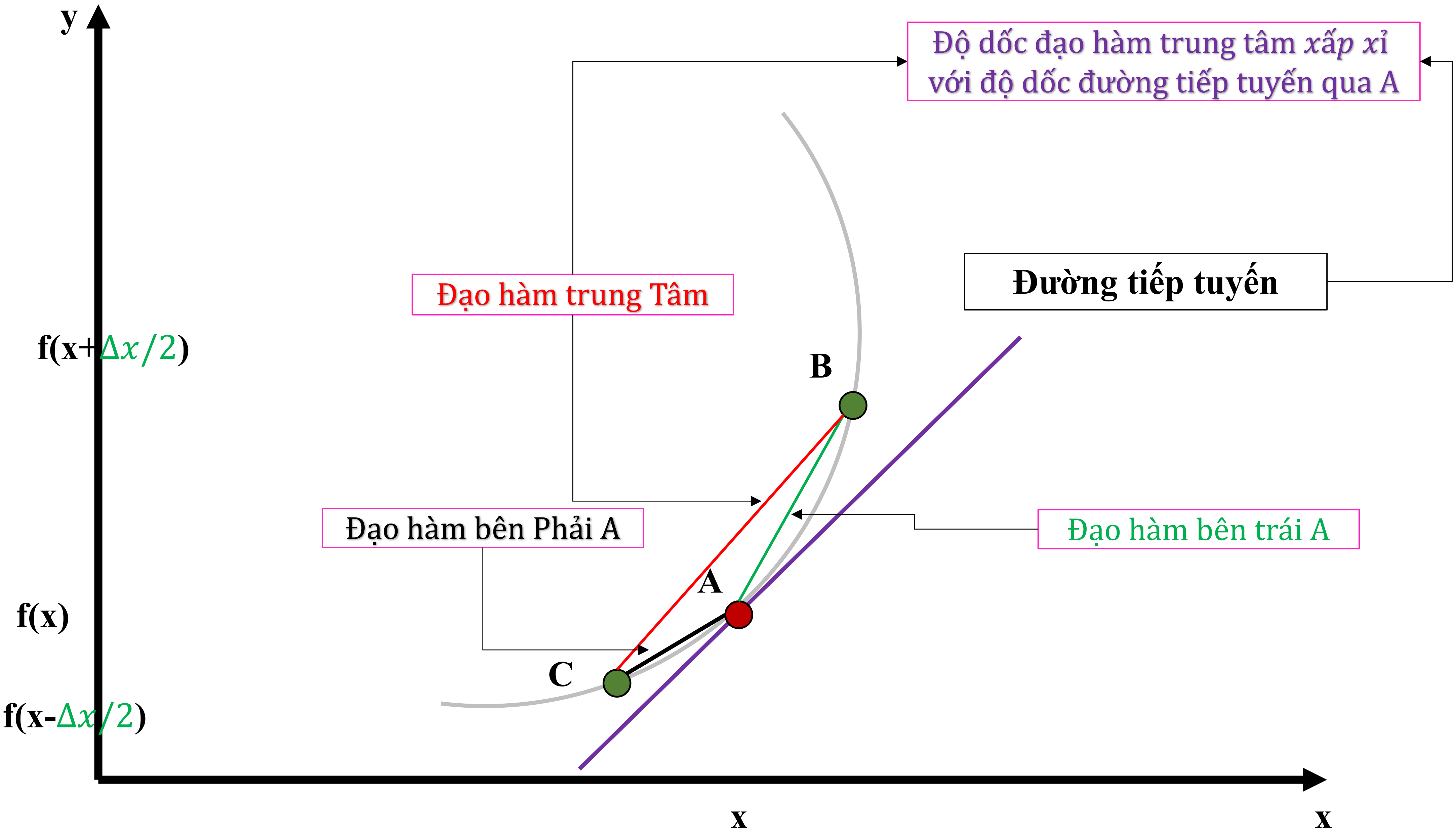


y
f(x+$\Delta x/2$)
B
Đường tiếp tuyến
A
f(x)
C
f(x-$\Delta x/2$)
$\Delta x/2$
$\Delta x/2$
x
x
Đạo hàm bên phải A $= \frac{d_y}{d_x} = \lim_{\Delta x \to 0} \frac{f(x+\Delta x/2) - f(x)}{\Delta x}$


y
f(x+Δx/2)
B
Đường tiếp tuyến
A
f(x)
C
f(x-Δx/2)
Δx/2
Δx/2
x
x
Đạo hàm bên trái A $= \frac{d_y}{d_x} = \lim_{\Delta x \to 0} \frac{f(x - \Delta x/2) - f(x)}{\Delta x}$

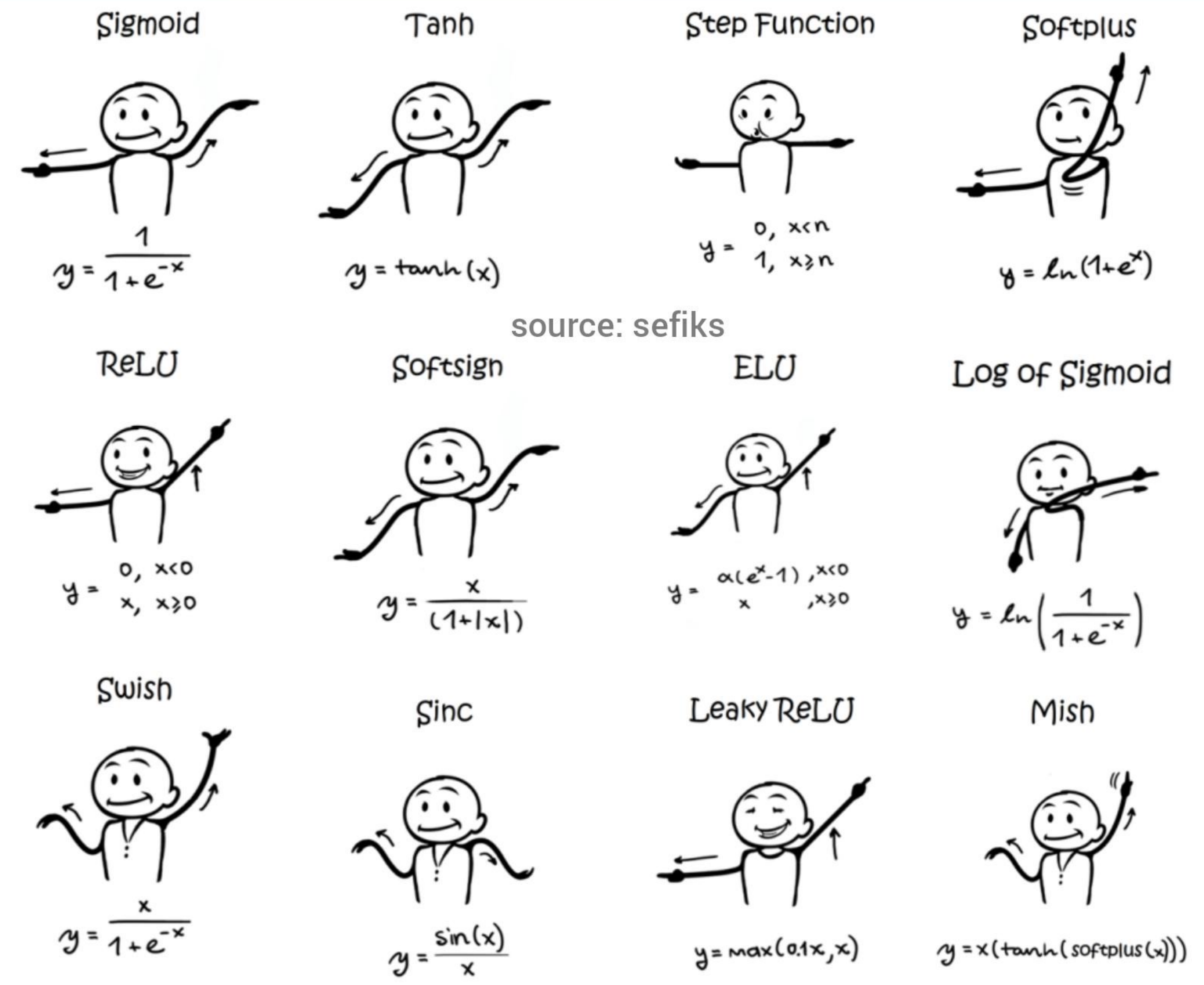


$$\text{Đạo hàm trung tâm A} = \frac{d_y}{d_x} = \lim_{\Delta x \to 0} \frac{f(x + \Delta x/2) - f\big((x - \Delta x/2)\big)}{\Delta x}$$



y
Độ dốc đạo hàm trung tâm *xấp xỉ* với độ dốc đường tiếp tuyến qua A
Đạo hàm trung Tâm
Đường tiếp tuyến
f(x+$\Delta x/2$)
B
Đạo hàm bên Phải A
Đạo hàm bên trái A
A
f(x)
C
f(x-$\Delta x/2$)
x
x

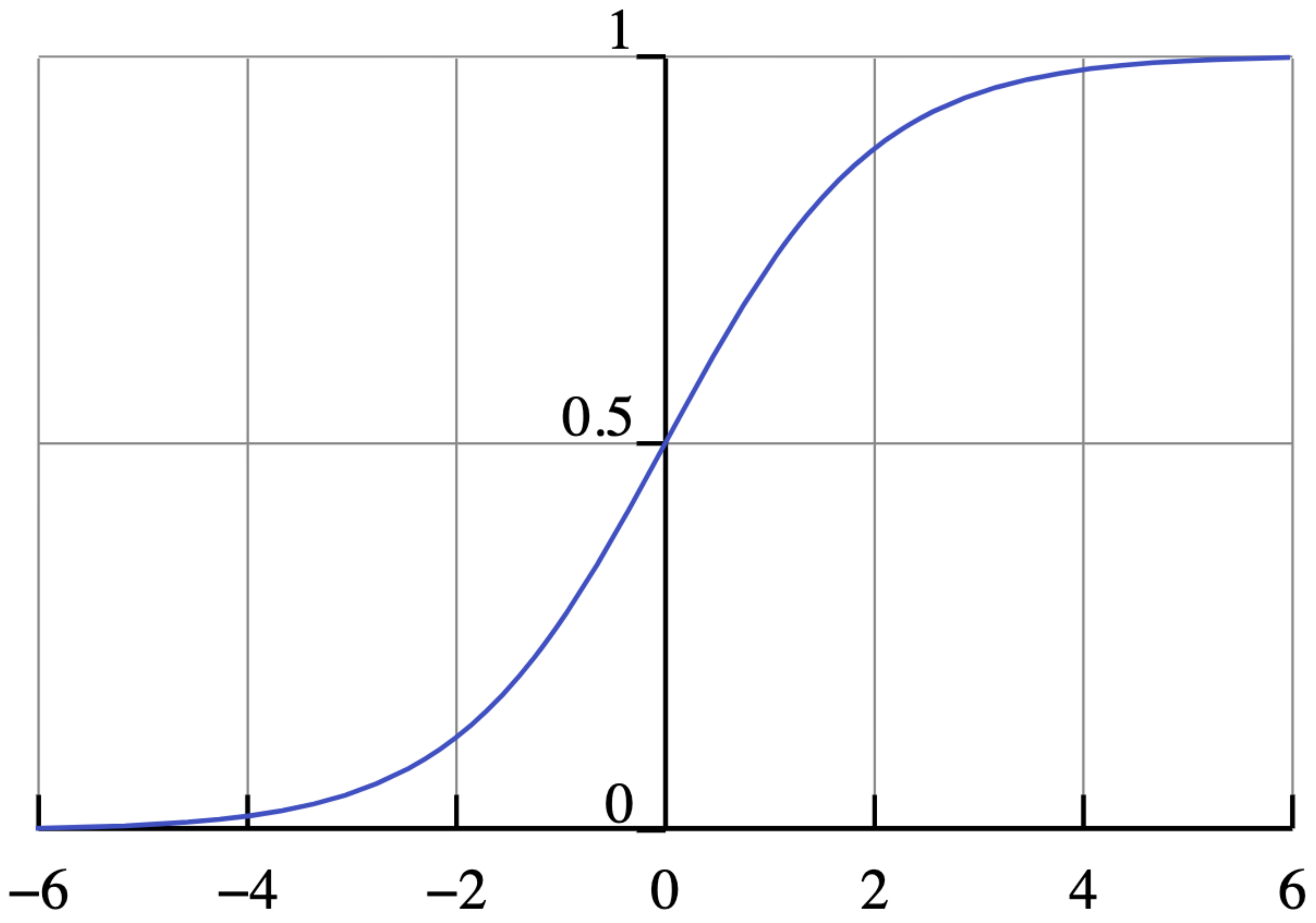
Dance Moves of Deep Learning Activation Functions



source: sefiks

What is an activation function and Why do we need to understand it?

$$sigmoid(x) = \frac{e^x}{e^x + 1}$$



Problem 1

1. **Trình bày chi tiết đạo hàm các activation sau và thực hiện code bằng python ((a) Tính đạo hàm và cài đặt; (b) dùng matplotlib để plot hàm số và hàm đạo hàm; chỉ dùng list)**

   1.1 **Exponential** $f(x) = e^x$

   1.2 **PReLU**: $\alpha = 0.25,\ f(x) = \begin{cases} \alpha x, & \text{if } x \leq 0. \\ x, & \text{if x} > 0. \end{cases}$

   1.3 **ELU**: $\alpha = 0.25,\ f(x) = \begin{cases} \alpha(e^x - 1), & \text{if } x \leq 0. \\ x, & \text{if } x > 0. \end{cases}$

   1.4 **Softplus**: $f(x) = \log(1 + e^x)$

   1.5 **Softsign**: $f(x) = \dfrac{x}{|x| + 1}$

Activation Functions

❖ Sigmoid function

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}}$$

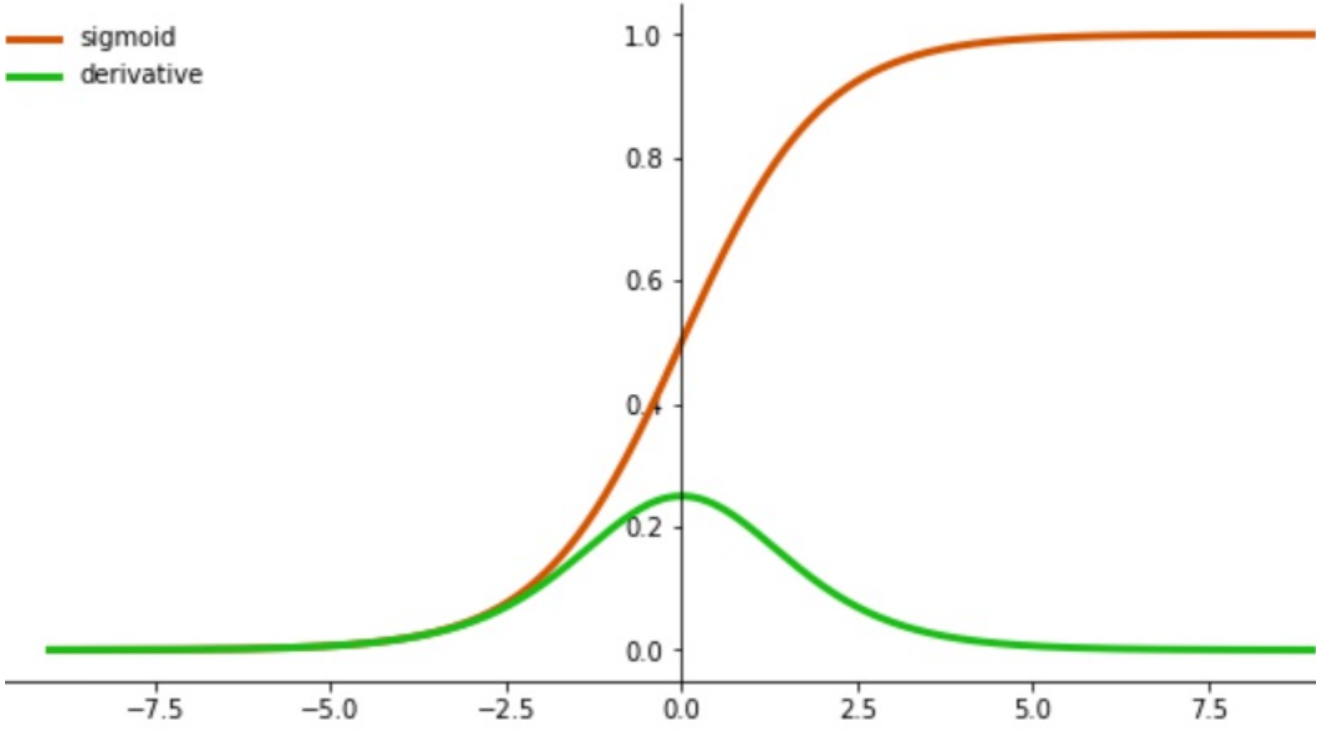
data =

1	5	-4	3	-2

data_a = sigmoid(data)

data_a =

0.731	0.993	0.017	0.95	0.119



$$\text{sigmoid}'(x) = \text{sigmoid}(x)\,(1 - \text{sigmoid}(x))$$

Activation Functions

$$\text{sigmoid}(x) = \frac{1}{1+e^{-x}}$$

$$\text{sigmoid}'(x) = \left(\frac{1}{1+e^{-x}}\right)' = \frac{-1}{(1+e^{-x})^2}(-e^{-x})$$

$$= \frac{e^{-x}}{(1+e^{-x})^2} = \frac{e^{-x}+1-1}{(1+e^{-x})^2}$$

$$= \frac{1}{1+e^{-x}} - \frac{1}{(1+e^{-x})^2}$$

$$= \frac{1}{1+e^{-x}}\left(1 - \frac{1}{1+e^{-x}}\right)$$

$$= \text{sigmoid}(x)\,(1 - \text{sigmoid}(x))$$

Activation Functions

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

$$\sigma'(x) = \frac{d}{dx}\sigma(x) = \frac{d}{dx}\frac{1}{1+e^{-x}}$$

$$= \frac{d}{dx}\frac{1}{1+e^{-x}} = \frac{d}{dx}(1+e^{-x})^{-1}$$

$$= \frac{d}{dx}(1+e^{-x})^{-1} = -(1+e^{-x})^{-2}.\frac{d}{dx}(1+e^{-x})$$

$$= -(1+e^{-x})^{-2}.(\frac{d}{dx}[1] + \frac{d}{dx}[e^{-x}])$$

$$= -(1+e^{-x})^{-2}.(e^{-x}.\frac{d}{dx}[-x])$$

$$= (1+e^{-x})^{-2}.e^{-x}$$

$$[\frac{1}{u(x)}]' = [u(x)^{-1}]' = -[\frac{u'(x)}{u(x)^2}] = -u(x)^{-2}.u'(x)$$

Activation Functions

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

$$\sigma'(x) = \frac{d}{dx}\sigma(x) = \frac{d}{dx}\frac{1}{1+e^{-x}}$$

$$= \frac{d}{dx}\frac{1}{1+e^{-x}} = \frac{d}{dx}(1+e^{-x})^{-1}$$

$$= \frac{d}{dx}(1+e^{-x})^{-1} = -(1+e^{-x})^{-2}.\frac{d}{dx}(1+e^{-x})$$

$$= -(1+e^{-x})^{-2}.(\frac{d}{dx}[1] + \frac{d}{dx}[e^{-x}])$$

$$= -(1+e^{-x})^{-2}.(e^{-x}.\frac{d}{dx}[-x])$$

$$= (1+e^{-x})^{-2}.e^{-x}$$

$$\text{sigmoid}'(x) = \left(\frac{1}{1+e^{-x}}\right)' = \frac{-1}{(1+e^{-x})^2}(-e^{-x})$$

$$= \frac{e^{-x}}{(1+e^{-x})^2} = \frac{e^{-x}+1-1}{(1+e^{-x})^2}$$

$$= \frac{1}{1+e^{-x}} - \frac{1}{(1+e^{-x})^2}$$

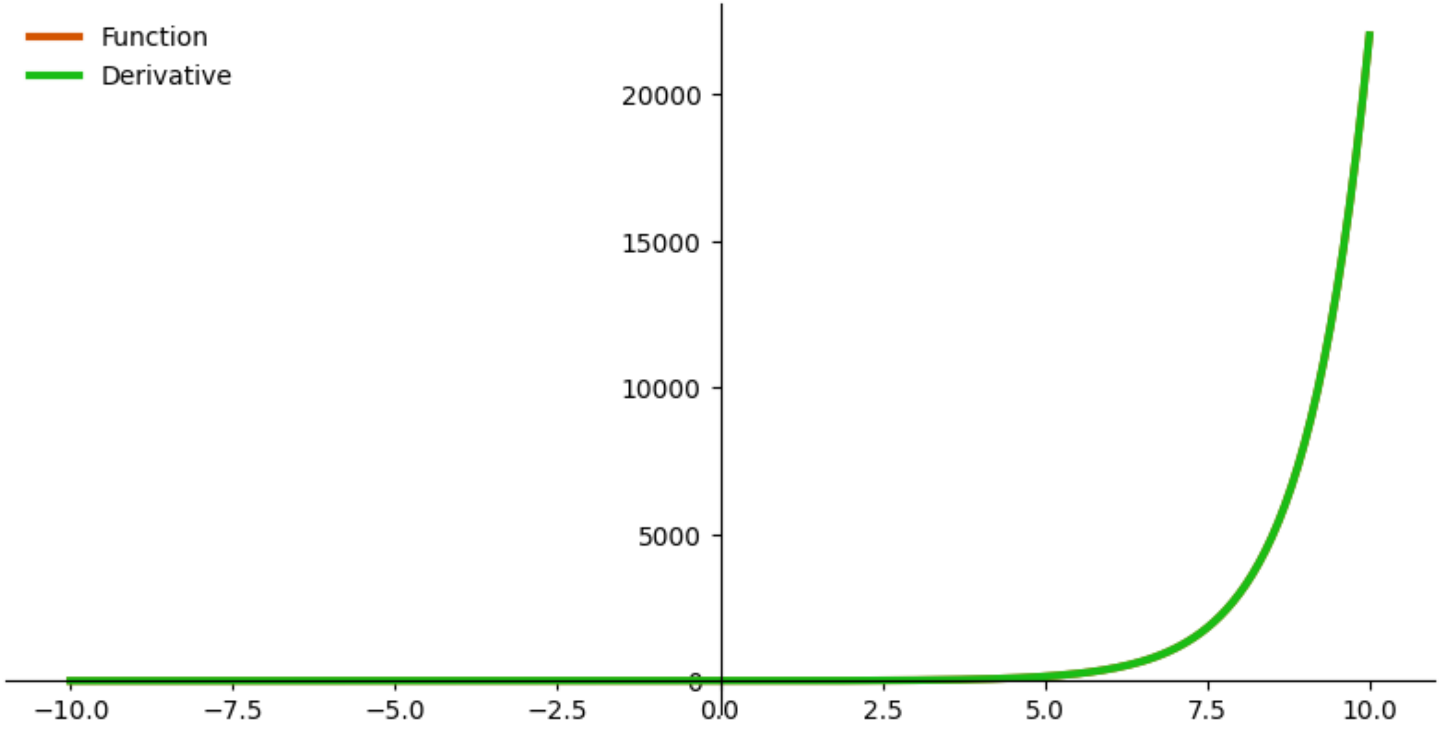
$$= \frac{1}{1+e^{-x}}\left(1 - \frac{1}{1+e^{-x}}\right)$$

$$= \text{sigmoid}(x)\,(1 - \text{sigmoid}(x))$$

Activation Functions

❖ Exponential function

- $f(x) = e^x$
- $f'(x) = e^x$



Activation Functions

❖ PReLU function

$$\text{PReLU}(x) = \begin{cases} \alpha x & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases}$$

$\alpha = 0.1$

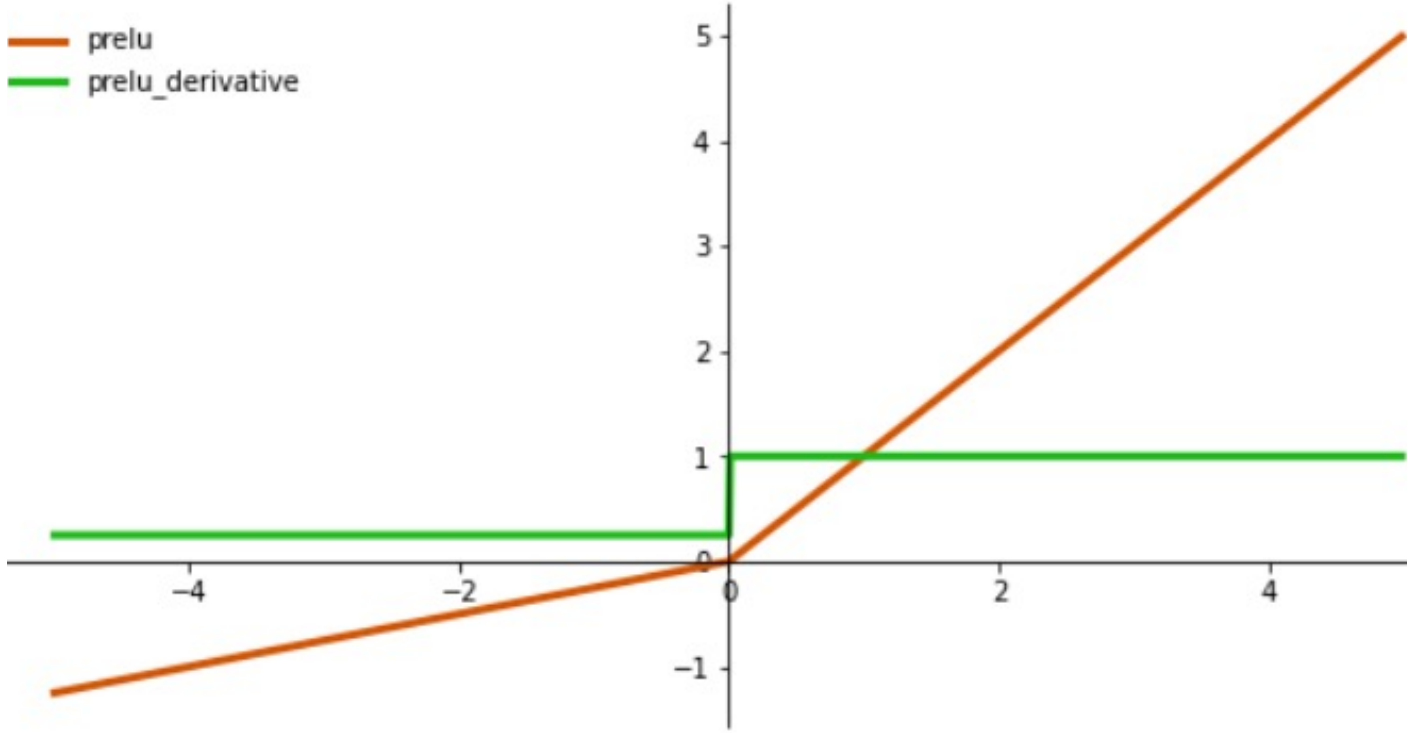
data =

1	5	-4	3	-2

data_a = PRELU(data)

data_a =

1	5	-0.4	3	-0.2



$$\text{PReLU}'(x) = \begin{cases} \alpha & \text{if } x \leq 0 \\ 1 & \text{if } x > 0 \end{cases}$$

Activation Functions

❖ ELU function

$$\text{ELU}(x) = \begin{cases} \alpha(e^x - 1) & \text{if } x \le 0 \\ x & \text{if } x > 0 \end{cases}$$

$$\alpha = 0.1$$

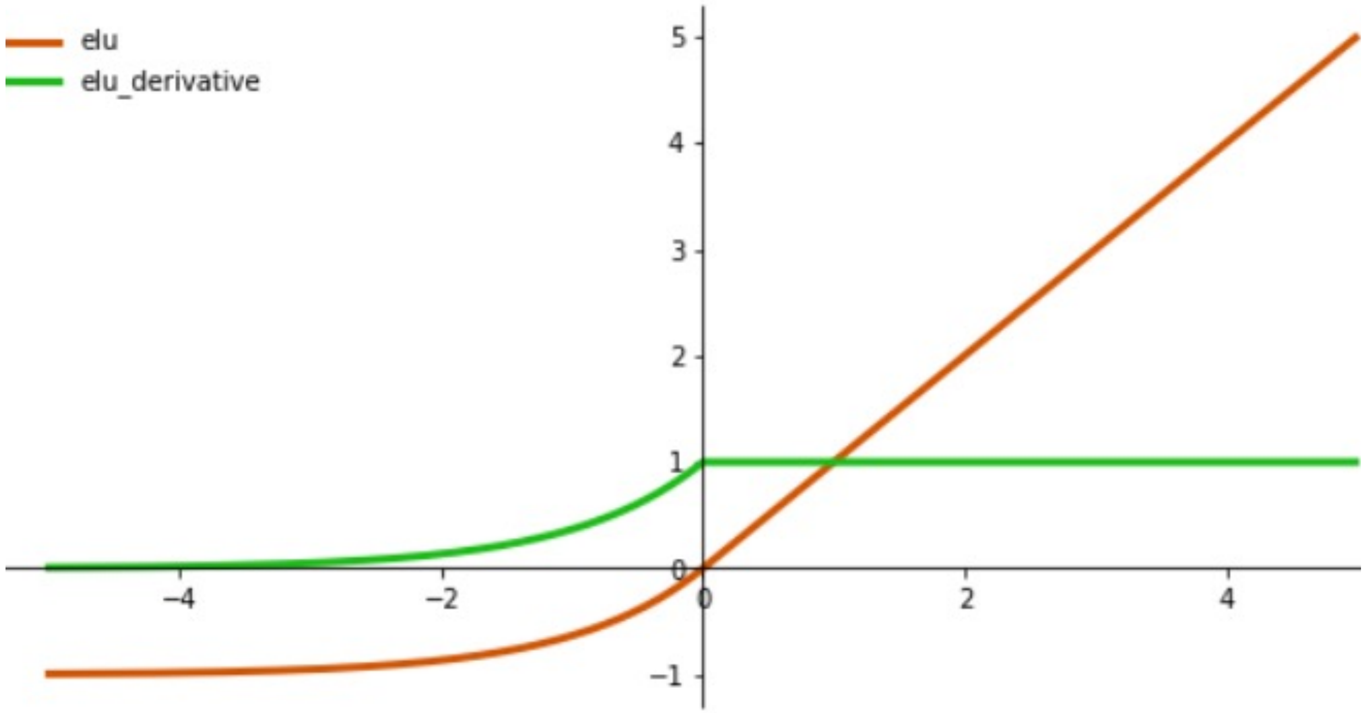
data =

1	5	-4	3	-2

data_a = ELU(data)

data_a =

1	5	-0.098	3	-0.086



$$\text{ELU}'(x) = \begin{cases} \alpha e^x & \text{if } x \le 0 \\ 1 & \text{if } x > 0 \end{cases}$$

Activation Functions

❖ Softplus function

$$\text{softplus}(x) = \log(1 + e^{x})$$

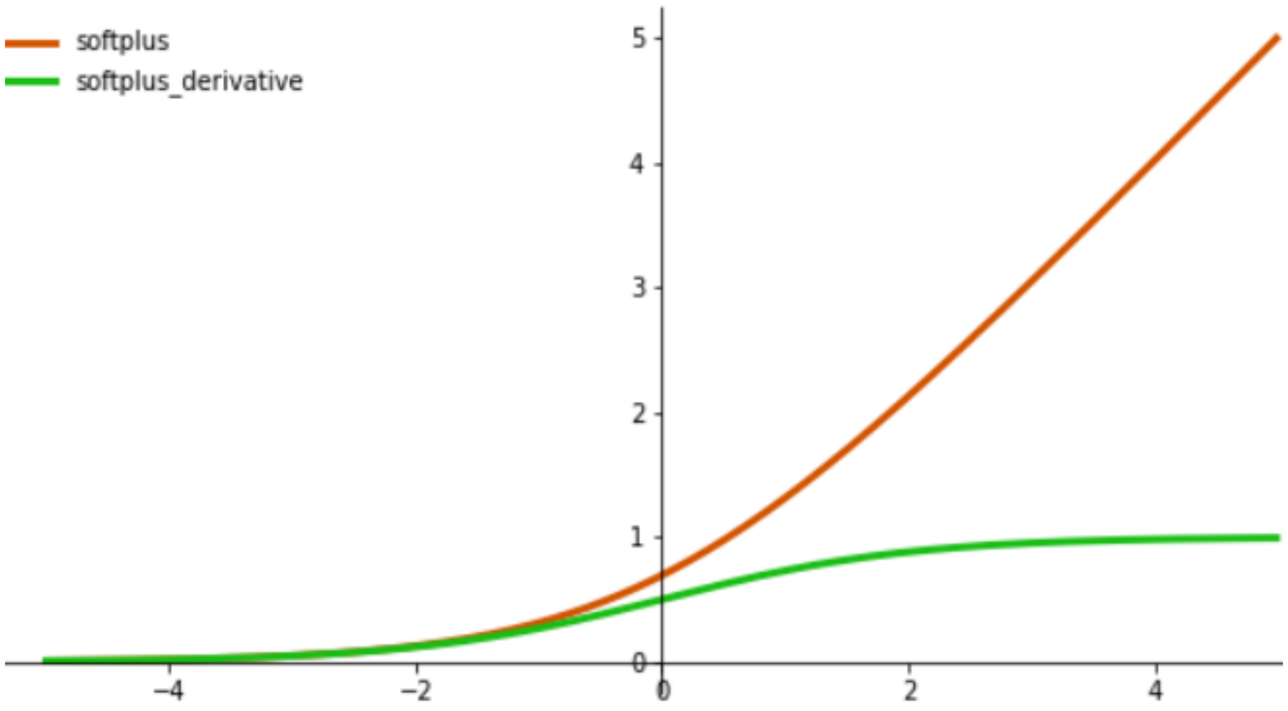
data =

1	5	-4	3	-2

data_a = softplus(data)

data_a =

1.313	5.006	0.018	3.048	0.126



$$\text{softplus}'(x) = \frac{1}{1 + e^{-x}}$$

$$\text{softplus}(x) = \log(1 + e^x)$$

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

$$\text{softplus}'(x) = (\log(1 + e^x))' = \left(\frac{1}{1 + e^x}\right)(e^x) = \frac{e^x}{1 + e^x} = \frac{\frac{e^x}{e^x}}{\frac{1}{e^x} + \frac{e^x}{e^x}}$$

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

Activation Functions

1.5 Softsign:

- $f(x) = \dfrac{x}{|x|+1}$

- $f'(x) = (\dfrac{x}{|x|+1})' = \dfrac{(x)'(|x|+1) - x(|x|+1)'}{(|x|+1)^2} = \dfrac{1(|x|+1) - x(\dfrac{x}{|x|})}{(|x|+1)^2}$

$$= \frac{|x|+1-(\dfrac{x^2}{|x|})}{(|x|+1)^2} = \frac{|x|\dfrac{|x|}{|x|}+1-(\dfrac{x^2}{|x|})}{(|x|+1)^2} = \frac{\dfrac{x^2}{|x|}+1-(\dfrac{x^2}{|x|})}{(|x|+1)^2}$$

$$= \frac{1}{(|x|+1)^2}$$

Possible derivation:

$$\frac{d}{dx}(|x|)$$

Rewrite $|x|$ as $\sqrt{x^2}$:

$$= \frac{d}{dx}\left(\sqrt{x^2}\right)$$

Using the chain rule, $\frac{d}{dx}\left(\sqrt{x^2}\right) = \frac{d\sqrt{u}}{du}\,\frac{du}{dx}$, where $u = x^2$ and $\frac{d}{du}\left(\sqrt{u}\right) = \frac{1}{2\sqrt{u}}$:

$$= \frac{\frac{d}{dx}(x^2)}{2\sqrt{x^2}}$$

Use the power rule, $\frac{d}{dx}(x^n) = n\,x^{n-1}$, where $n = 2$.

$\frac{d}{dx}(x^2) = 2\,x$:

$$= \boxed{2\,x}\,\frac{1}{2\sqrt{x^2}}$$

Simplify the expression:

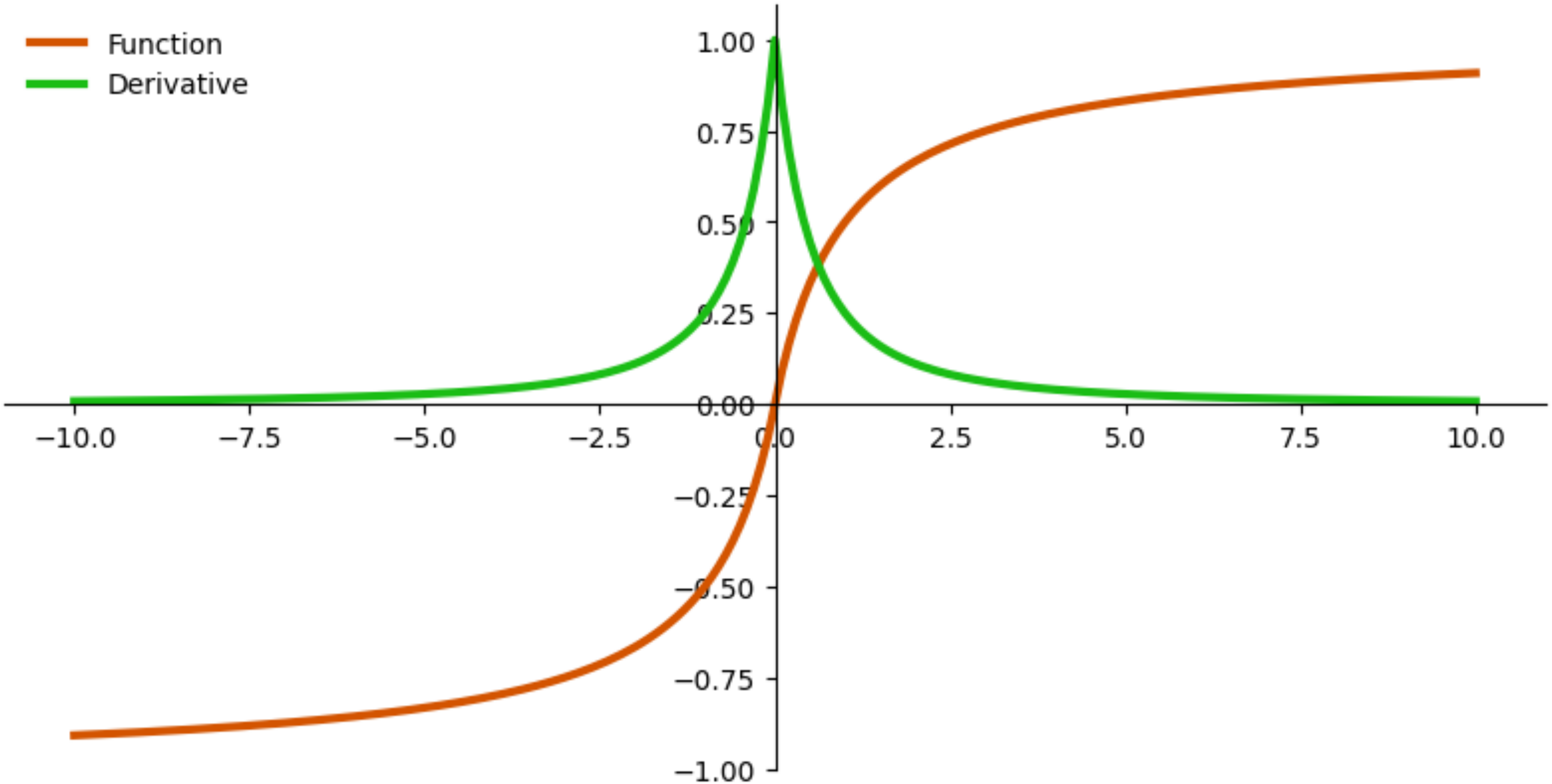
$$= \frac{x}{\sqrt{x^2}}$$

Rewrite $\frac{1}{\sqrt{x^2}}$ as $\frac{1}{|x|}$:

Answer:

$$= \frac{x}{|x|}$$

❖ Softsign function



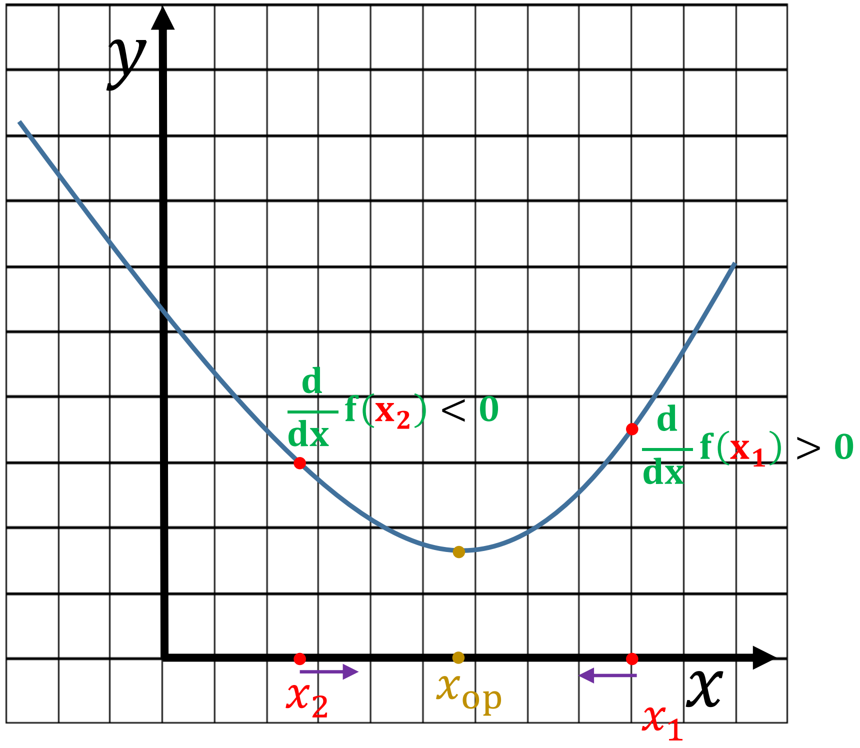
Problem 2

2.1 Viết function *find_minimum(f, x, num_iteration, step)* và dựa theo thuật toán ở hình 2 để tìm xấp xỉ x mà $f(x)$ (ví dụ $f(x) = 3x^4 - 4x^2 - 6x - 3$) là minimum.

- **Input**: Nhận 4 input
  - **f**: function $f(x)$
  - **x**: giá trị khởi tạo x đầu tiên
  - **num_iteration**: Số lần lặp thuật toán để tìm x
  - **step**: độ lớn để cho một lần cập nhật x (độ lớn quãng đường đi ngược hướng với giá trị đạo hàm tại x (dx))
- **Output** giá trị xấp xỉ của x tại đó giá trị hàm $f(x)$ (hàm được truyền vào từ input) là minimum
- **Các bạn thực hiện theo các step sau**:
  - **Step1**: Thực hiện vòng lặp với num_iteration số lần lặp
  - **Step2**: Trong mỗi lần lặp tìm giá trị đạo hàm (dx) tại x của hàm $f(x)$ bằng phương pháp đạo hàm trung tâm (central difference)
  - **Step3**: Xét dấu của giá trị đạo hàm (dx) để xác định độ lớn giá trị cập nhật.
  - **Step4**: Nếu dx là số dương thì cập nhật $x = x - step$, nếu dx là số âm thì cập nhật $x = x + step$, nếu $dx = 0$ thì không thực hiện việc cập nhật
  - **Step5**: Thực hiện Step2, Step3 và Step4 để cập nhật x cho đến khi đủ num_iteration số lần lặp thì thoát loop và trả về x mới nhất

Gradient-based Optimization

Tìm giá trị min



Quan sát: x_{op} **ở vị trí ngược hướng đạo hàm tại** $\mathbf{x_1}$ **và** $\mathbf{x_2}$

Cách xử lý việc di chuyển ngược hướng đạo hàm cho $\mathbf{x_1}$ **và** $\mathbf{x_2}$ **(để tìm** x_{op}**) khác nhau hình thành các thuật toán tối ưu hóa khác nhau**

Cách cập nhật giá trị x đơn giản

$$\mathbf{x} = \mathbf{x} - \eta \frac{\mathbf{d}}{\mathbf{dx}}\mathbf{f(x)}$$

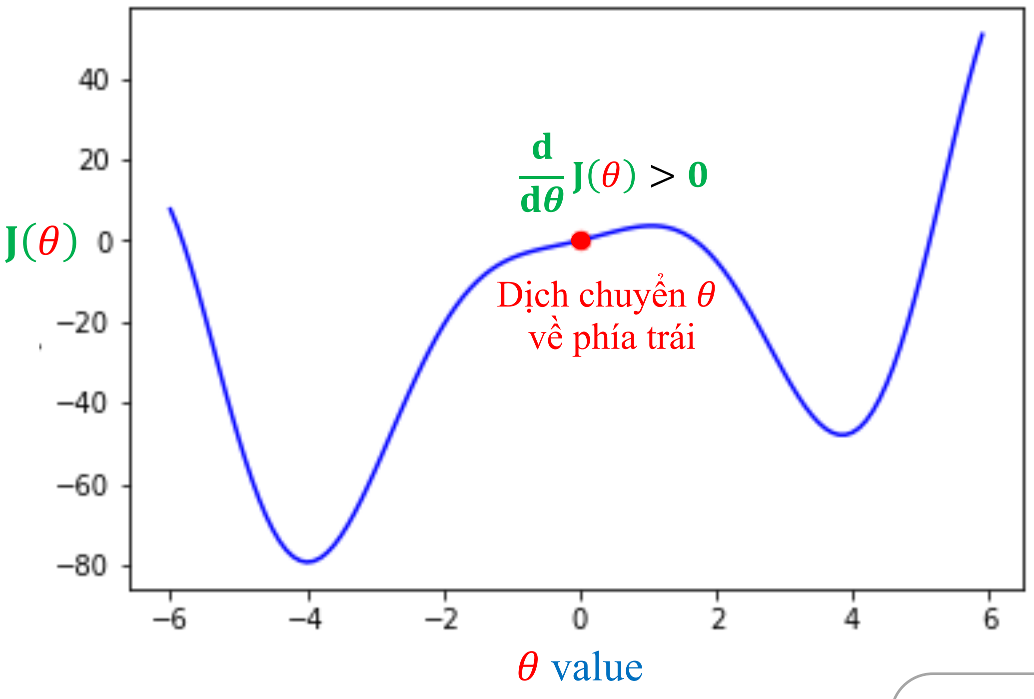
Đạo hàm tại x

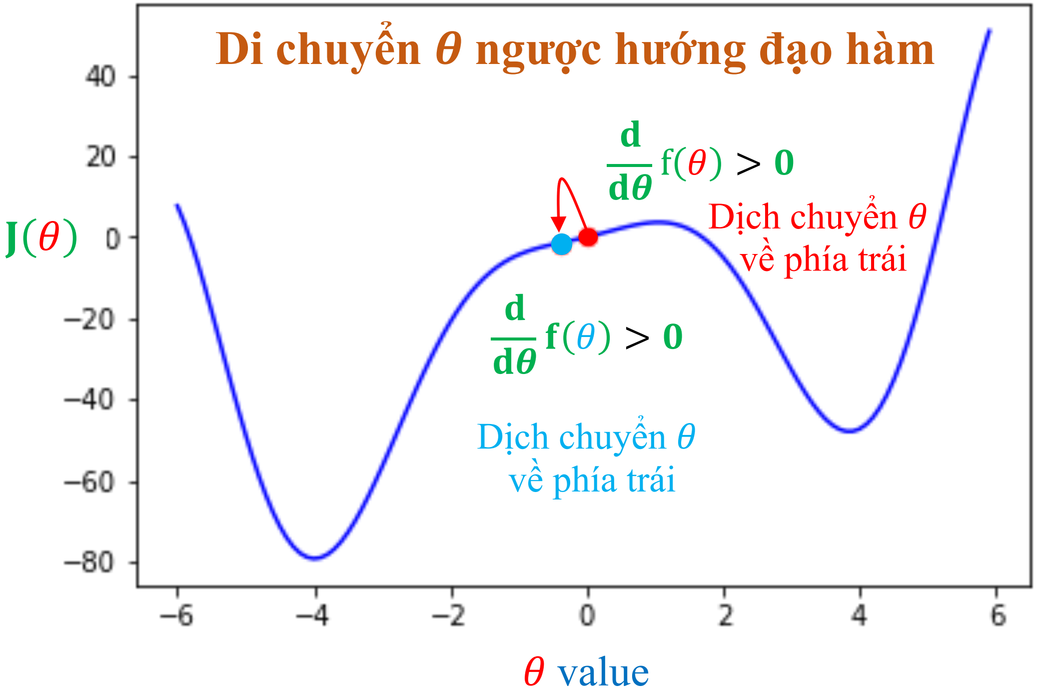
Trọng số

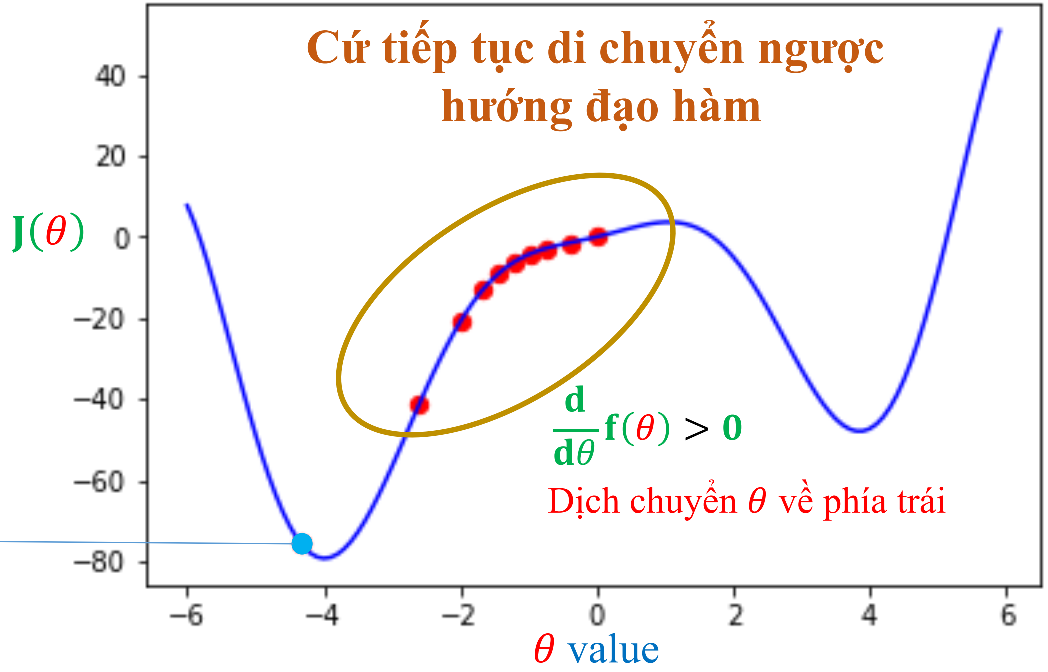
Optimization

❖ **A cue to optimize a function**

Khởi tạo giá trị θ







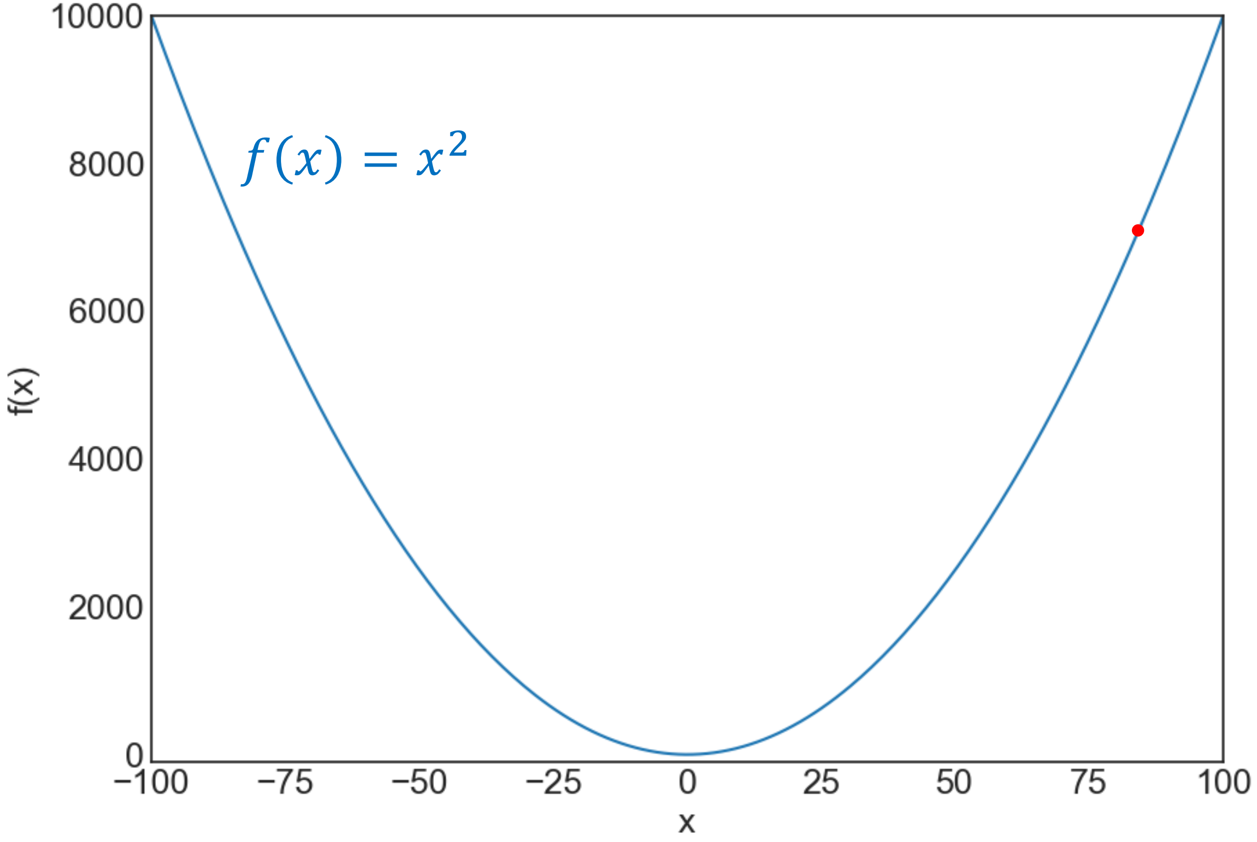
$\frac{d}{d\theta}f(\theta) < 0$

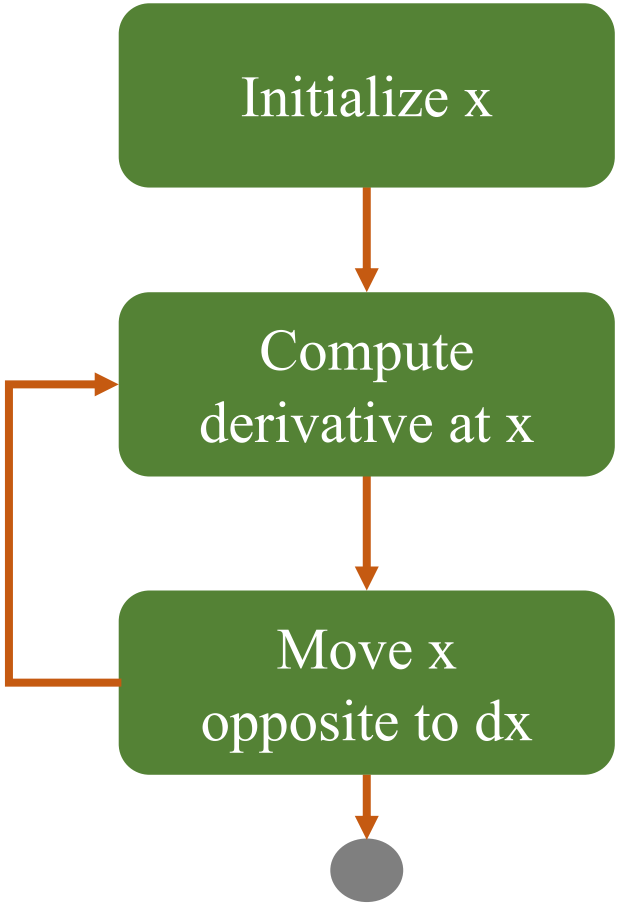
Dịch chuyển θ
về phía phải

Gradient-based Optimization

❖ Square function

$$x_t = x_{t-1} - \eta f'(x_{t-1})$$





$$-100 \leq x \leq 100$$

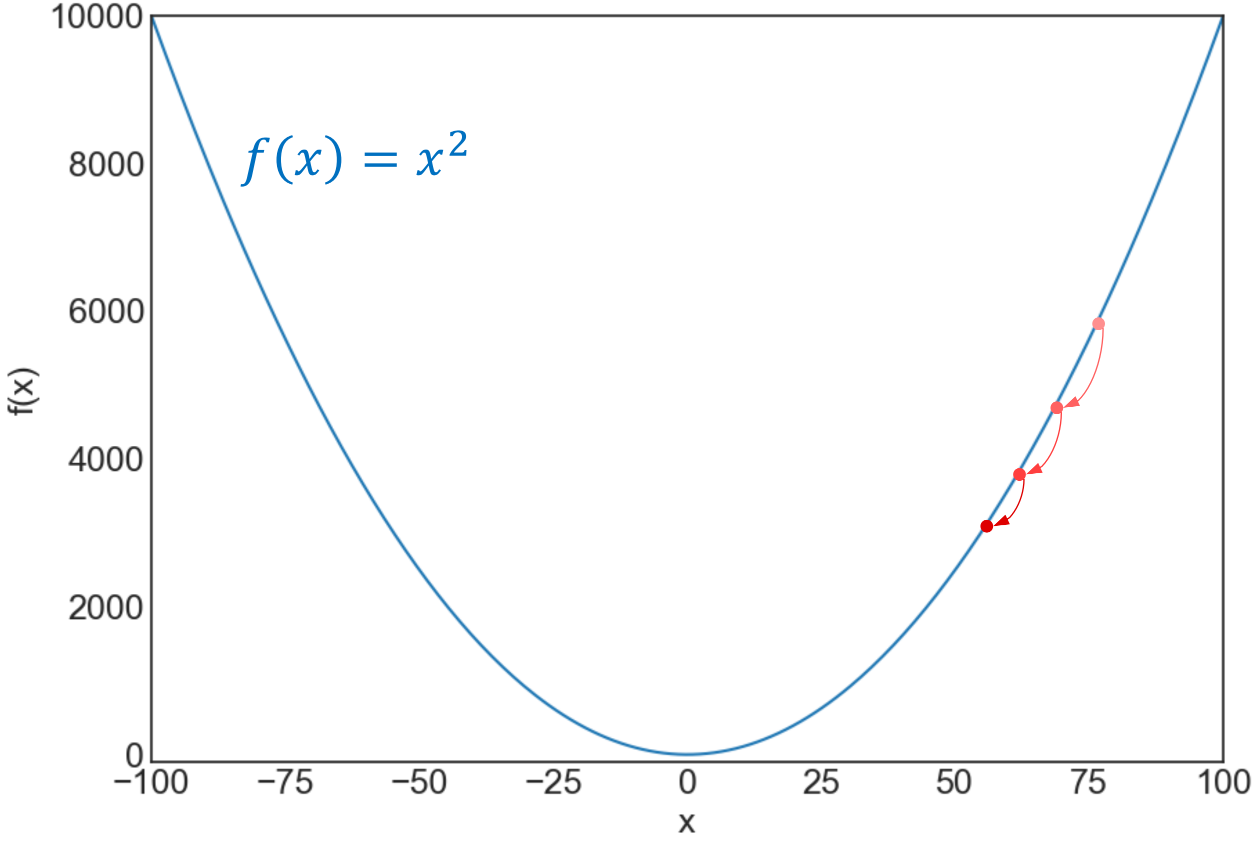
$$x \in \mathbb{N}$$

AI VIETNAM
All-in-One Course

Optimization

$$x_t = x_{t-1} - \eta f'(x_{t-1})$$

❖ Square function



$$-100 \le x \le 100$$
$$x \in \mathbb{N}$$

$x_0 = 70.0$ $\quad \eta = 0.1$

$f'(x_0) = 140.0$

$x_1 = x_0 - \eta f'(x_0) = 56.0$

$f'(x_1) = 112.0$

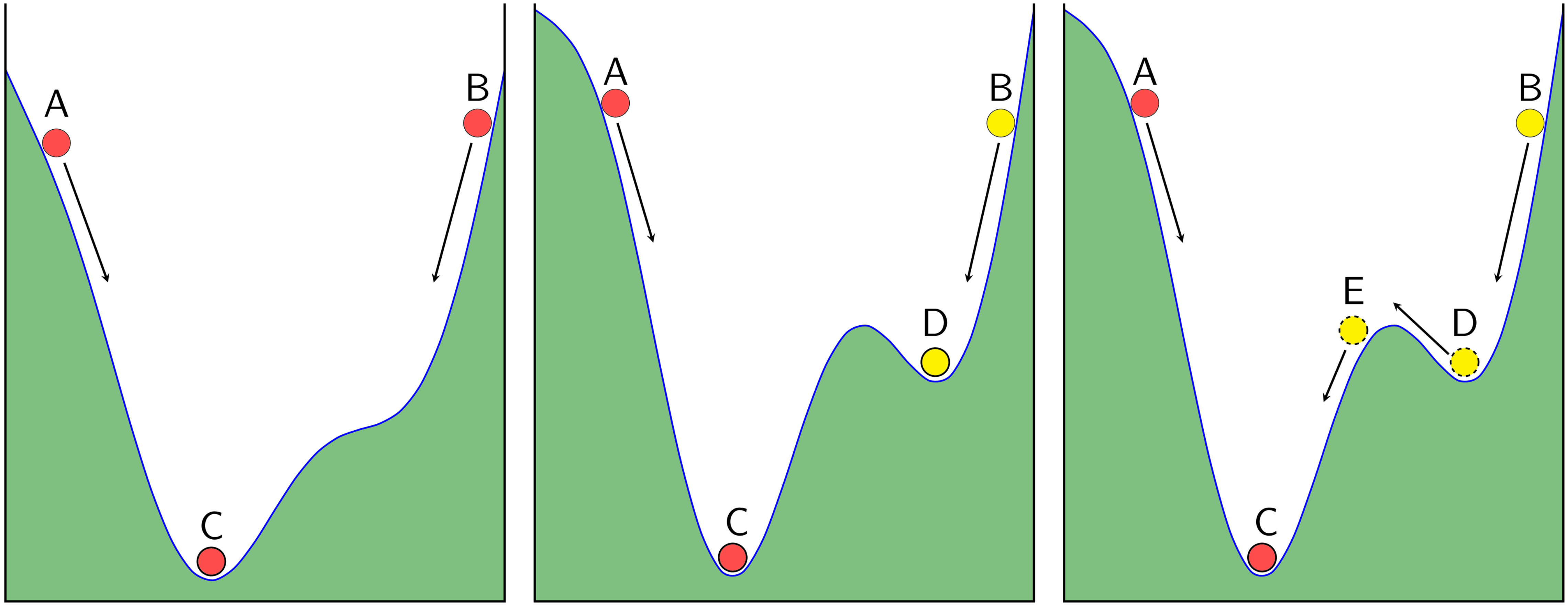
$x_2 = x_1 - \eta f'(x_1) = 44.8$

$f'(x_2) = 89.6$

$x_3 = x_2 - \eta f'(x_2) = 35.84$

$f'(x_3) = 71.68$

$x_4 = x_3 - \eta f'(x_3) = 28.672$



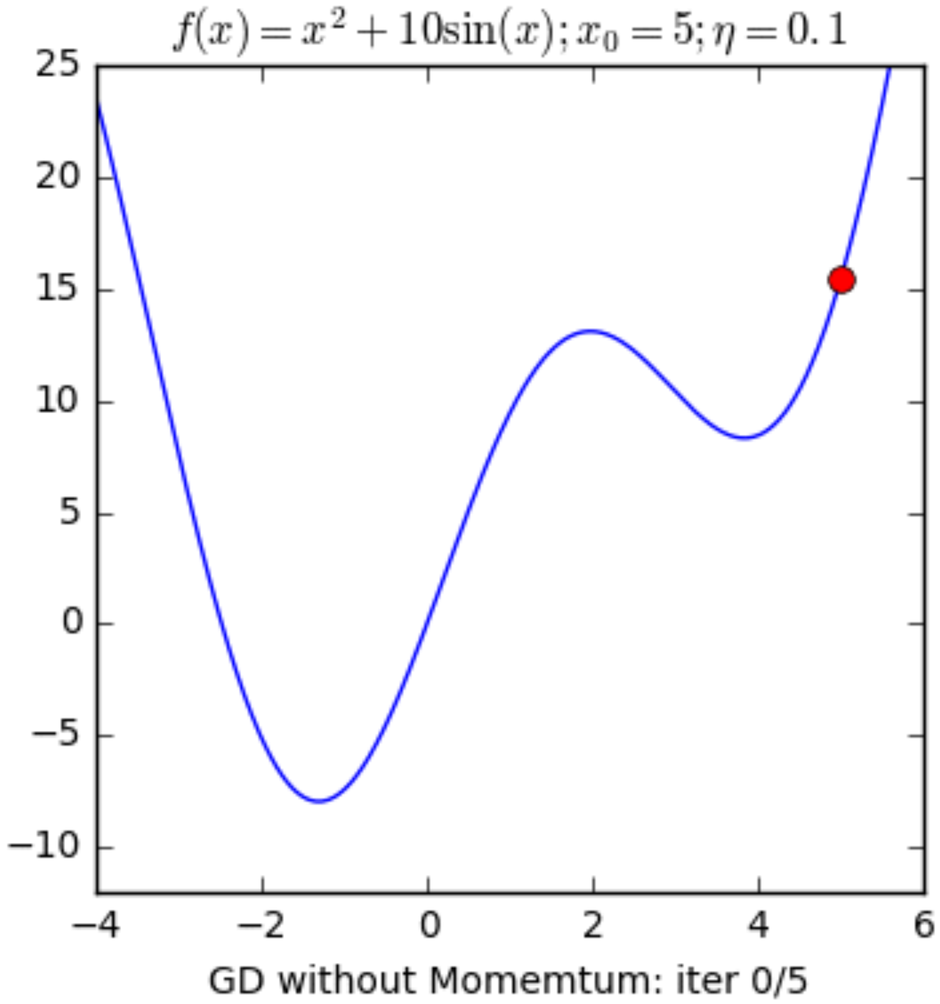
a) GD

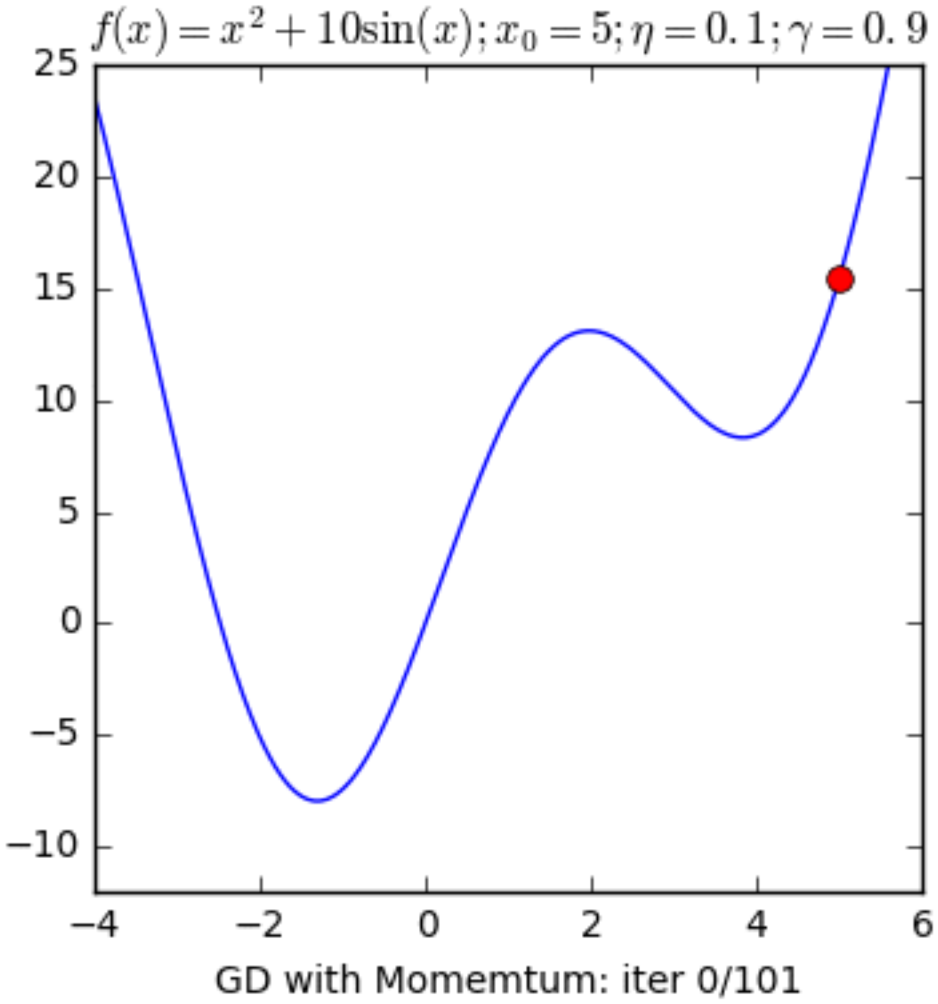
b) GD

c) GD with momentum

$$x_t = x_{t-1} - \eta f'(x)$$

$$v_t = \gamma v_{t-1} - \eta \frac{d}{dx} f(x)$$
$$x_t = x_{t-1} - v_t$$


$f(x) = x^2 + 10\sin(x); x_0 = 5; \eta = 0.1$
25
20
15
10
5
0
−5
−10
−4
−2
0
2
4
6
GD without Momemtum: iter 0/5


$f(x) = x^2 + 10\sin(x); x_0 = 5; \eta = 0.1; \gamma = 0.9$
25
20
15
10
5
0
−5
−10
−4
−2
0
2
4
6
GD with Momemtum: iter 0/101

Problem 3

3. **Sử dụng đạo hàm rời rạc trên ảnh để phát hiện thông tin cạnh/edge (optional), như hình 3**. Cho trước một ảnh đầu vào LINK có kích thước bất kỳ. Hãy viết chương trình giảm size ảnh xuống (400,400), sau đó hiện thức các hàm sau:

```python
def computeXDerivative(image):
  w = len(image[0])
  h = len(image)
  x_derivative = [[0]*w for _ in range(h)]

  # your code here *************

  return x_derivative
```

- *computeYDerivative(image))*:

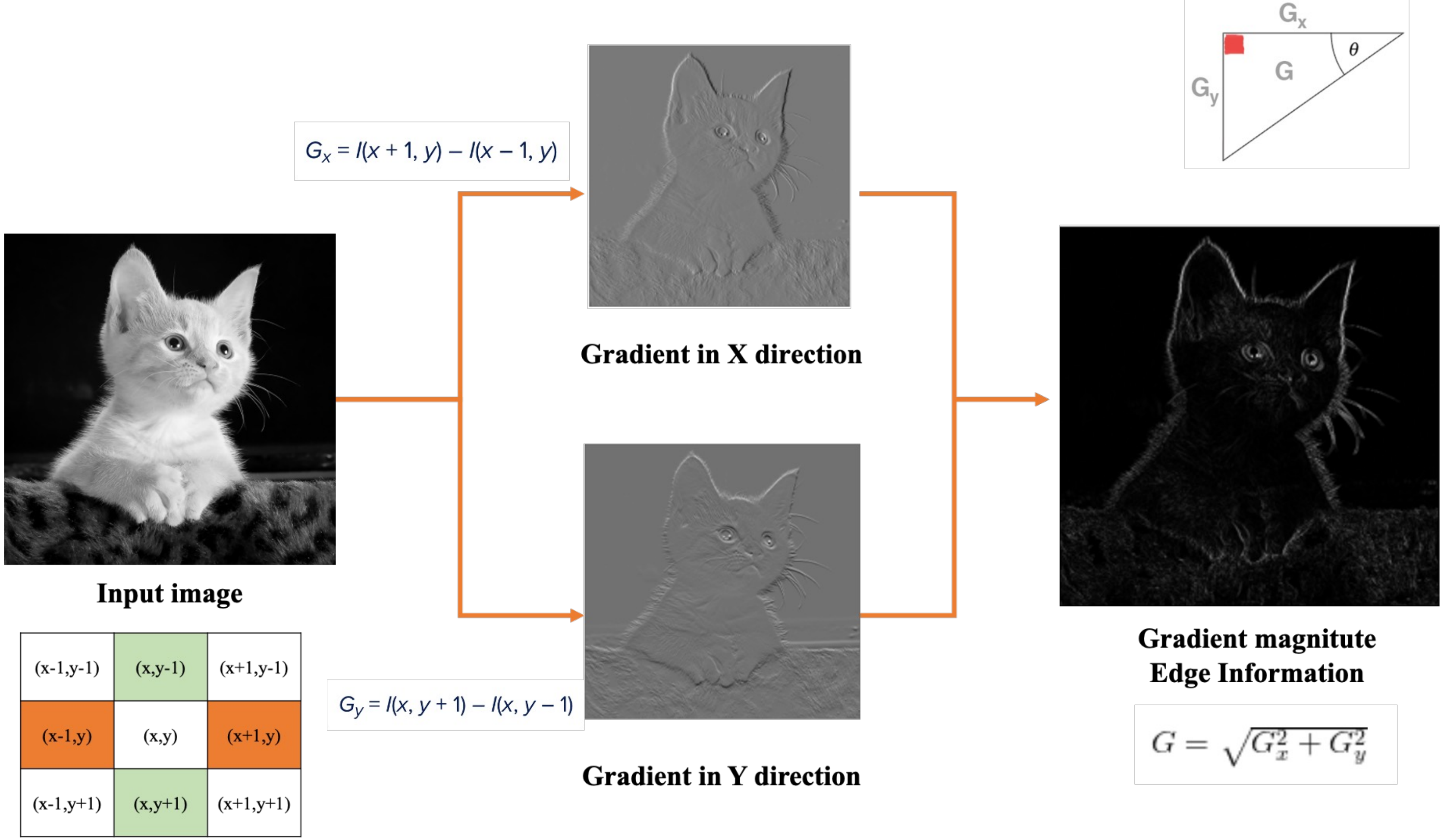
```python
#Caculate gradient in Y-direction
def computeYDerivative(image):
  w = len(image[0])
  h = len(image)
  y_derivative = [[0]*w for _ in range(h)]

   # your code here *************

  return y_derivative
```

- *computeMagniguteXY(image))*:

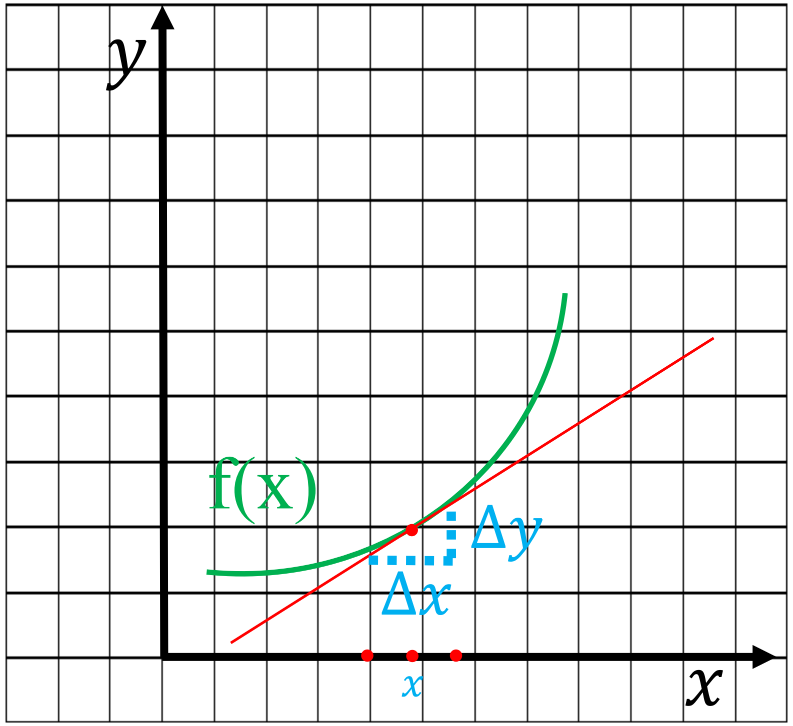
```python
# Calculate the gradient magnitude
def computeMagniguteXY(image):
  w = len(image[0])
  h = len(image)
  gradient_magnitude = [[0]*w for _ in range(h)]

   # your code here *************

  return gradient_magnitude
```

Problem 3

3. **Sử dụng đạo hàm rời rạc trên ảnh để phát hiện thông tin cạnh/edge (optional), như hình 3.** Cho trước một ảnh đầu vào LINK có kích thước bất kỳ. Hãy viết chương trình giảm size ảnh xuống (400,400), sau đó hiện thức các hàm sau:



Áp dụng cho hàm rời rạc



Đạo hàm$= \dfrac{\text{Thay đổi theo } y}{\text{Thay đổi theo } x} = \dfrac{\Delta y}{\Delta x}$

$$\frac{d}{dx} f(x) = \lim_{\Delta x \to 0} \frac{f\left(x + \frac{\Delta x}{2}\right) - f\left(x - \frac{\Delta x}{2}\right)}{\Delta x}$$

f(x) =

32	30	45	36	160	156	155	170

x

$\Delta x = 2$

$$\frac{d}{dx} f(x) = \frac{f(x+1) - f(x-1)}{2} = \frac{156-36}{2} = 60$$

-1	0	1

x-derivative filter

Derivative and Applications

Tính đạo hàm trung bình theo hướng x

$$\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} * \begin{bmatrix} -1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

weighted average — x-derivative — Sobel for x direction

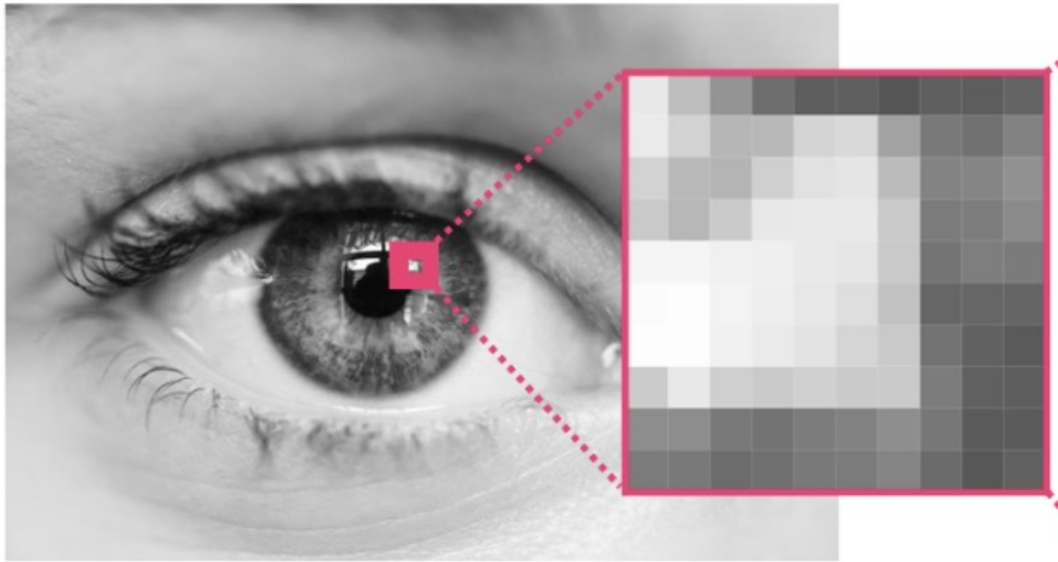
Tính đạo hàm trung bình theo hướng y

$$\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} * \begin{bmatrix} 1 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

y-derivative — weighted average — Sobel for y direction

Derivative and Applications

❖ Grayscale images



230	194	147	108	90	98	84	96	91	101
237	206	188	195	207	213	163	123	116	128
210	183	180	205	224	234	188	122	134	147
198	189	201	227	229	232	200	125	127	135
249	241	237	244	232	226	202	116	125	126
251	254	241	239	230	217	196	102	103	99
243	255	240	231	227	214	203	116	95	91
204	231	208	200	207	201	200	121	95	95
144	140	120	115	125	127	143	118	92	91
121	121	108	109	122	121	134	106	86	97

(Height, Width)

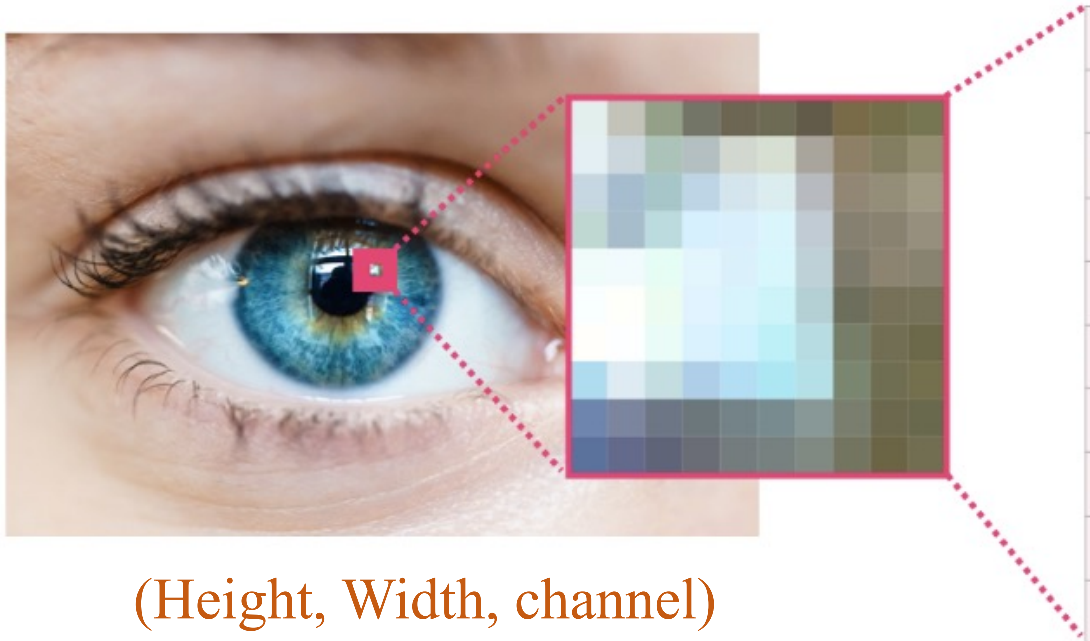
Pixel p = scalar

$0 \leq p \leq 255$

Resolution: #pixels
Resolution = Height x Width

Derivative and Applications

❖ Color images



226	191	147	110	101	112	98	123	110	119
221	191	176	182	203	214	169	144	133	145
185	160	161	184	205	223	186	137	147	161
181	174	189	207	206	215	194	136	142	151
246	237	237	231	208	206	192	122	143	144
254	254	241	224	199	192	181	99	122	117
239	248	232	207	187	182	184	110	114	110
193	215	193	167	158	164	181	114	112	111
113	119	110	111	113	123	135	120	108	106
93	97	91	103	107	111	122	112	104	114

(Height, Width, channel)

RGB color image

Pixel p= $\begin{bmatrix} r \\ g \\ b \end{bmatrix}$

$0 \leq r,g,b \leq 255$

Resolution: #pixels
Resolution = Height x Width

Derivative and Applications

❖ Grayscale images

$$a = -230 + 147 - 2\times 237 + 2\times 188 - 210 + 180 = -211$$

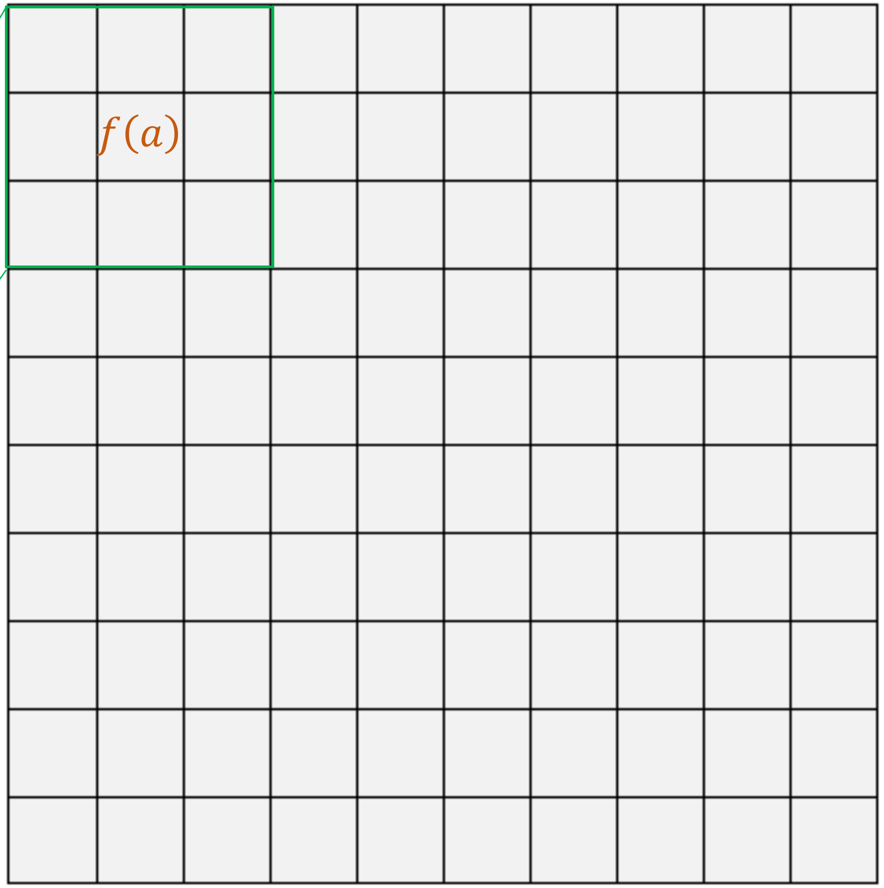
$$f(a) = |a| = 211$$

230	194	147	108	90	98	84	96	91	101
237	206	188	195	207	213	163	123	116	128
210	183	180	205	224	234	188	122	134	147
198	189	201	227	229	232	200	125	127	135
249	241	237	244	232	226	202	116	125	126
251	254	241	239	230	217	196	102	103	99
243	255	240	231	227	214	203	116	95	91
204	231	208	200	207	201	200	121	95	95
144	140	120	115	125	127	143	118	92	91
121	121	108	109	122	121	134	106	86	97

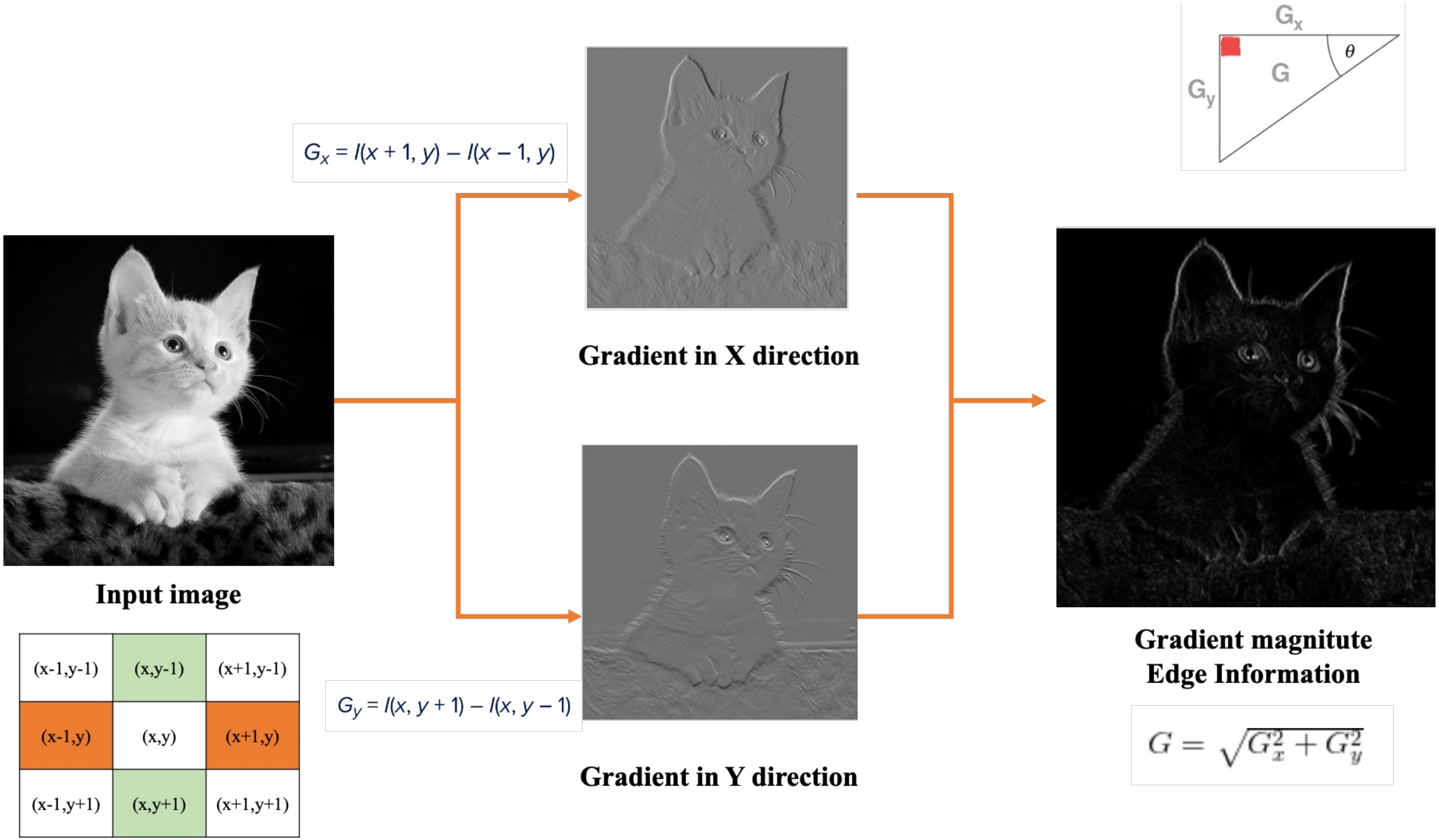
Input Image

-1	0	1
-2	0	2
-1	0	1

Sobel for x direction

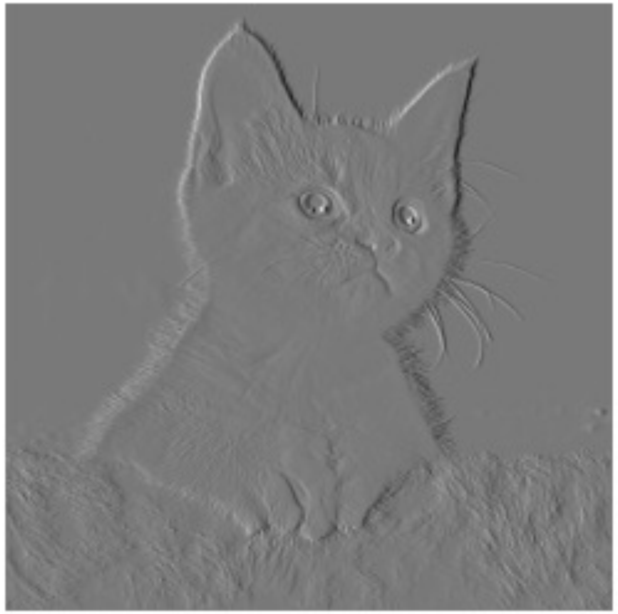


IMPLEMENT IMAGE GRADIENT WITH AND WITHOUT USING COVOLUTION

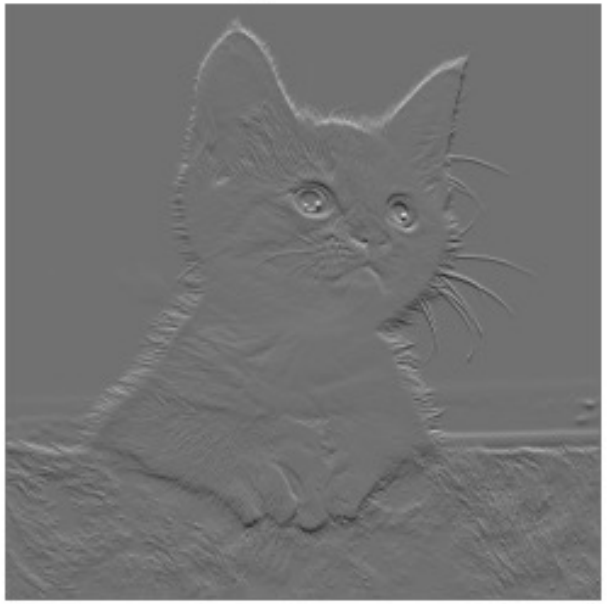


G_x
θ
G
G_y
$G_x = I(x+1, y) - I(x-1, y)$
Gradient in X direction
Input image
(x-1,y-1)
(x,y-1)
(x+1,y-1)
(x-1,y)
(x,y)
(x+1,y)
(x-1,y+1)
(x,y+1)
(x+1,y+1)
$G_y = I(x, y+1) - I(x, y-1)$
Gradient in Y direction
Gradient magnitute
Edge Information
$G = \sqrt{G_x^2 + G_y^2}$

```python
#Caculate gradient in X-direction
def computeXDerivative(image):
  w = len(image[0])
  h = len(image)
  x_derivative = [[0]*w for _ in range(h)]

  for i in range(1,h-1,1):
    for j in range(1,w-1,1):
      x_derivative[i][j] = image[i][j+1] -  image[i][j-1]

  return x_derivative
```



```python
#Caculate gradient in Y-direction
def computeYDerivative(image):
  w = len(image[0])
  h = len(image)
  y_derivative = [[0]*w for _ in range(h)]

  for i in range(1,h-1,1):
    for j in range(1,w-1,1):
      y_derivative[i][j] = image[i+1][j] -  image[i-1][j]

  return y_derivative
```



```
[11] # Calculate the gradient magnitude
     def computeMagnigute(image):
       x_d = computeXDerivative(image)
       y_d = computeYDerivative(image)

       w = len(image[0])
       h = len(image)

       gradient_magnitude = [[0]*w for _ in range(h)]

       for i in range(1,h-1,1):
         for j in range(1,w-1,1):
           gradient_magnitude[i][j] = math.sqrt((x_d[i][j])**2 + (y_d[i][j])**2)


       return gradient_magnitude
```

Derivative and Convolution

❖ Grayscale images

$$a = -230 + 147 - 2\times 237 + 2\times 188 - 210 + 180 = -211$$

$$f(a) = |a| = 211$$

230	194	147	108	90	98	84	96	91	101
237	206	188	195	207	213	163	123	116	128
210	183	180	205	224	234	188	122	134	147
198	189	201	227	229	232	200	125	127	135
249	241	237	244	232	226	202	116	125	126
251	254	241	239	230	217	196	102	103	99
243	255	240	231	227	214	203	116	95	91
204	231	208	200	207	201	200	121	95	95
144	140	120	115	125	127	143	118	92	91
121	121	108	109	122	121	134	106	86	97

Input Image

-1	0	1
-2	0	2
-1	0	1

Sobel for x direction

	$f(a)$								

```python
#Caculate gradient in X-direction
def computeGradientUsingConv(image, kernel):
  w = len(image[0])
  h = len(image)
  x_derivative = [[0]*w for _ in range(h)]
  k = 1
  for i in range(1,h-1,1):
    for j in range(1,w-1,1):
      image_region = (row[j-k: j+k+1] for row in cat_image_list[i-k: i+k+1])
      x_derivative[i][j] = sum( sum(j[0] * j[1] for j in zip(i[0], i[1])) for i in zip(kernel, image_region))
  return x_derivative
```

```python
x_kernel = [[0, 0, 0], [-1, 0, 1], [0, 0, 0]]
x_gradient = computeGradientUsingConv(cat_image_list, x_kernel)

y_kernel = [[0, -1, 0], [0, 0, 0], [0, 1, 0]]

y_gradient = computeGradientUsingConv(cat_image_list, y_kernel)
```