수포자도 도전해 볼 만한

Mathematics in DeepLearning

Lecture3. Multivariate Calculus

Juhee Lee ph.D.
Research Professor @ ewha womans university

Preliminary

Definition A function f is a rule that assigns to each element x in a set X exactly one element, called $f(x)$ in a set Y.

$$f : X \to Y$$
$$x \longmapsto f(x)$$

Usually,

sets X, Y : sets of real numbers

X : the domain, Y : the codomain of the function

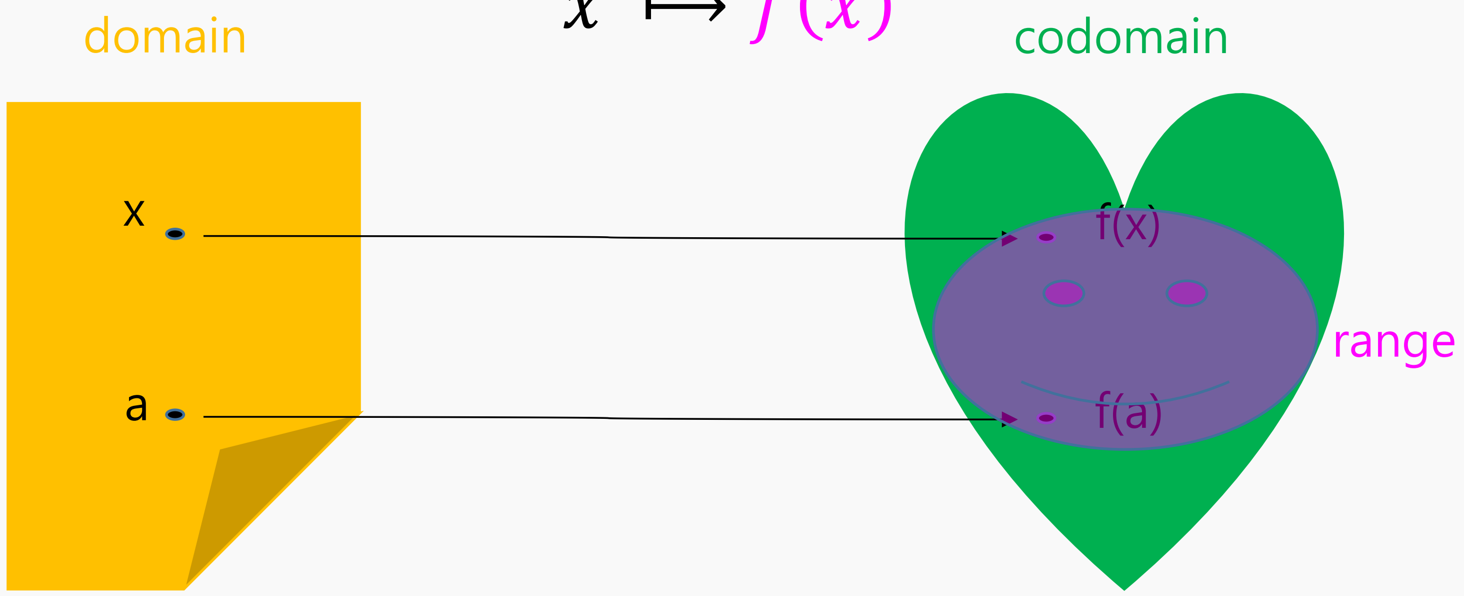
$f(x)$: the value of f at x

The range of f : the set of all possible values of $f(x)$

x : independent variable, $f(x)$: dependent variable

A function

$$f : X \rightarrow Y$$
$$x \longmapsto f(x)$$



injective function (one to one function)

surjective function (onto)

bijective function (one to one + onto)

Linear model

$$y = ax + b$$

Polynomials

$$P(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

Power function

$$\mathrm{f}(x) = x^a$$

Rational function

$$f(x) = P(x)/Q(x)$$

Trigonometric function

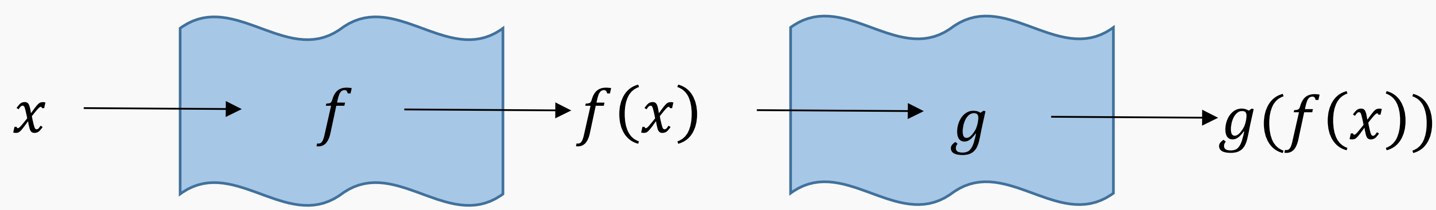
$$f(x) = \sin x$$

Exponential function and Logarithm

$$f(x) = x^a \text{ and } f(x) = \log_a x$$

<u>Definition</u> Given two functions f and g, the **composite function** $g \circ f$ is defined by

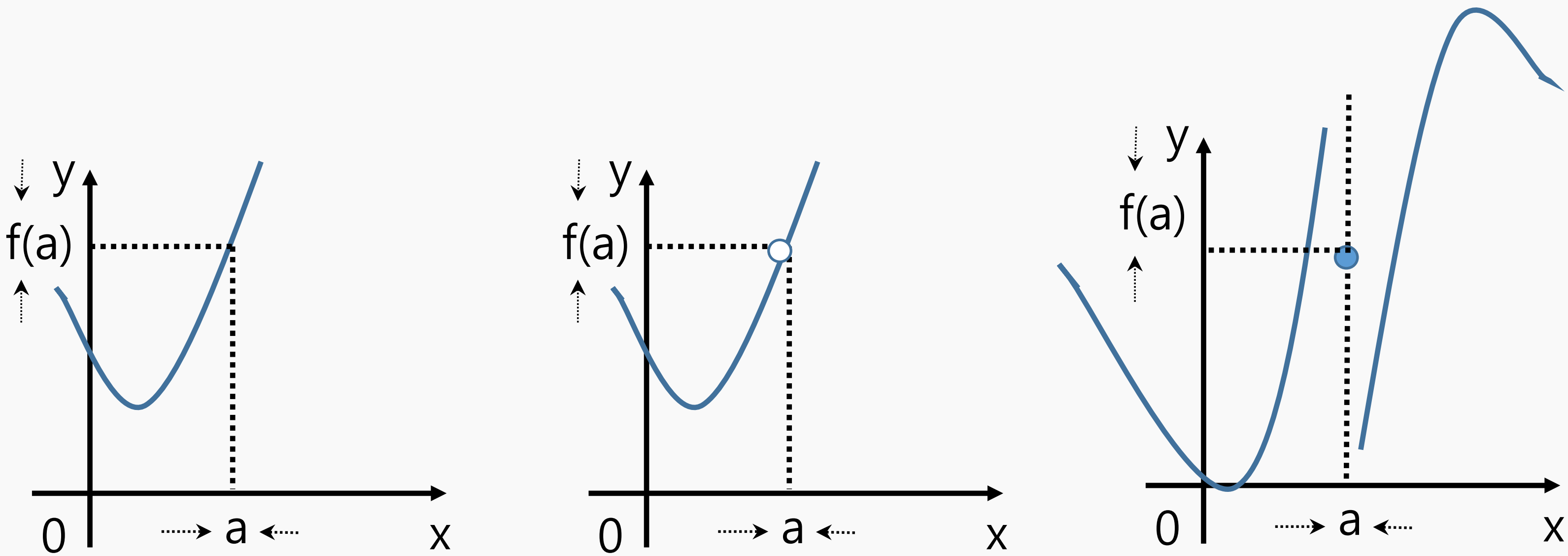
$$(g \circ f)(x) = g(f(x))$$



$$\frac{1}{2}$$

$$\frac{1}{2}+\left(\frac{1}{2}\right)^2$$

$$\frac{1}{2}+\left(\frac{1}{2}\right)^2+\left(\frac{1}{2}\right)^3+\cdots$$



As x approaches a

$$\lim_{x \to a} f(x) = L \Leftrightarrow \lim_{x \to a^-} f(x) = L \text{ and } \lim_{x \to a^+} f(x) = L$$

ex. Find $\lim_{x \to 0} \frac{1}{x^2}$ if it exists.

Definition The limit of $f(x)$,

as x approaches a, equals L

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

$$\Leftrightarrow \forall \epsilon > 0, \exists \delta > 0 \ s.t$$
$$0 < |x - a| < \delta \ \Rightarrow |f(x) - L| < \epsilon$$

ex. Show that $\lim_{x \to 0} |x| = 0$.

ex. Show that $\lim_{x \to 0} \frac{|x|}{x}$ does not exists.

Definition A function f is continuous at a number a if

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

A function f is **continuous on an interval** I if it is continuous at every number in the interval

$$\lim_{x \to a} f(x) = f(a),\ a \in I$$

Definition A function f is continuous at a number a if

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

정리하면,

step1. $f(a)$: $f(x)$가 $x = a$에서 정의되고

step2. $\lim_{x \to a} f(x)$가 존재하고

step3. $\lim_{x \to a} f(x) = f(a)$ 이면

$x = a$에서 연속이다.

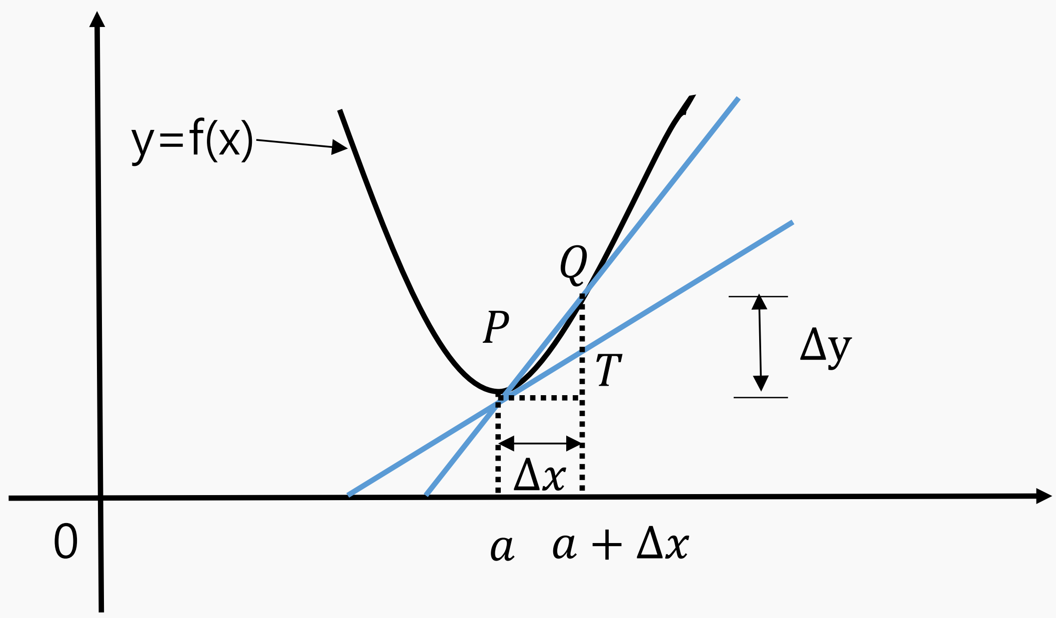
Definition A function f is defined on some open interval that continuous at a number a, except possibly at a itself. Then

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

means that for every positive number M there is a positive number δ such that

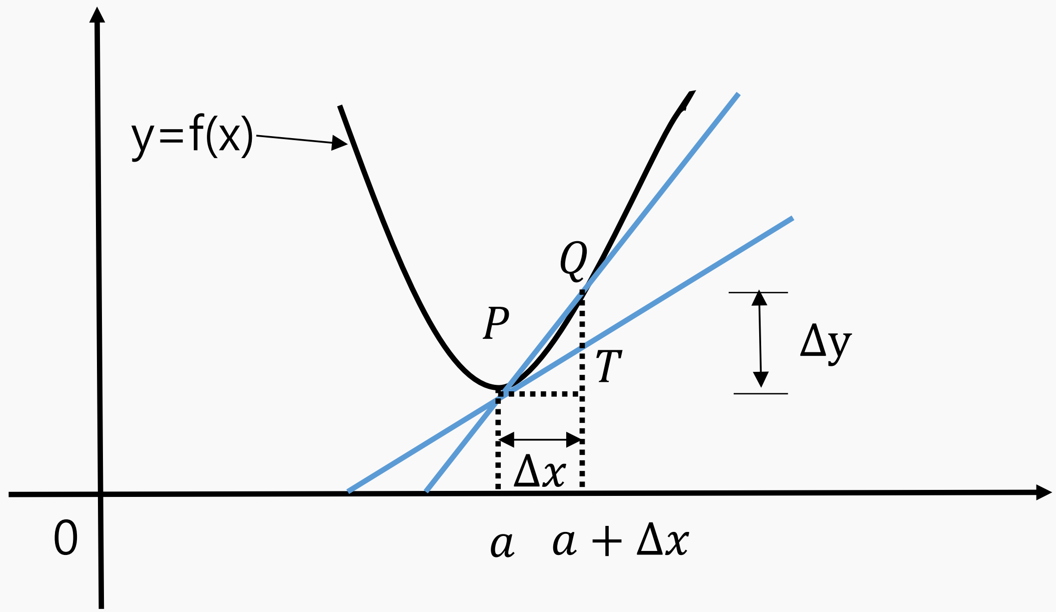
If $0 < |x - a| < \delta$ then $f(x) > M$

Instantaneous rate of change

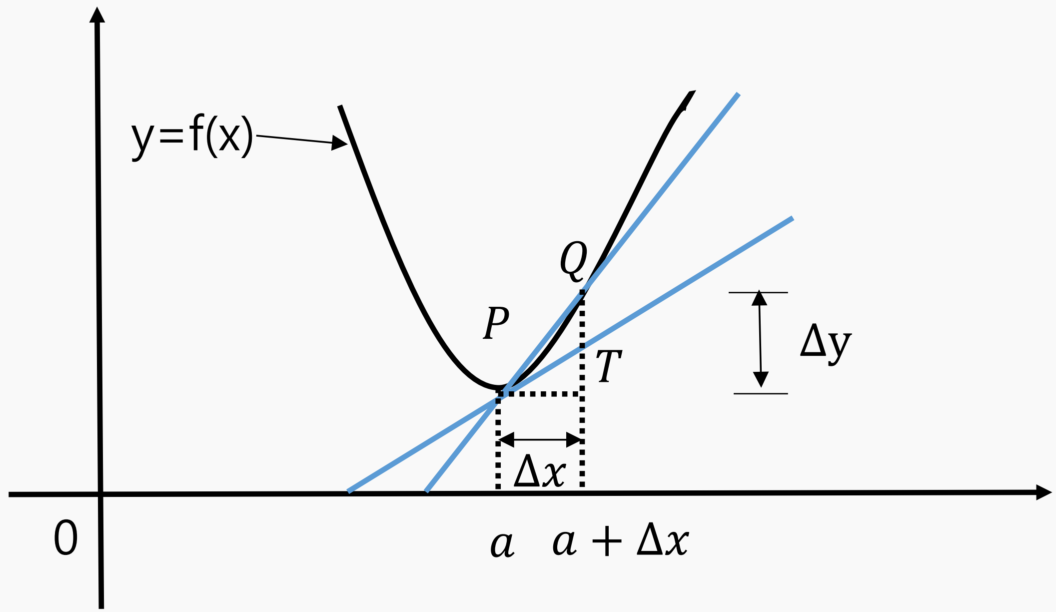


$$\frac{\Delta y}{\Delta x} = \frac{f(a + \Delta x) - f(a)}{\Delta x}$$

Differential coefficient



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



$\exists f'(a) \Rightarrow \Delta x \rightarrow 0$ 일때, $\mathrm{Q} \rightarrow P$ 이고 직선 $P\mathrm{Q}$는 기울기가 $f'(a)$인 직선 PT로 접근한다.

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

$$= \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

$$\forall a \in X \text{ , } \Delta \mathrm{x} = h = x_2 - x_1$$

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

$$= \lim_{h \to 0} \frac{f(x+h) - f(x-h)}{2h}$$

$$= \lim_{x_2 \to x_1} \frac{f(x_2) - f(x_1)}{x_2 - x_1}$$

y = f(x) 의 도함수 $f'(x)$가 존재하고, $f'(x)$의 도함수가 존재하면, 이것을 처음 함수 f(x)의 이계 도함수,

$$f''(\mathrm{x}) \text{ or } y'' \text{ or } \frac{d^2y}{dx^2} \text{ or } \frac{d^2}{dx^2}f(x) \text{ or } D^2f(x) \text{ or } D^2y$$

$$f^{(n)}(\mathrm{x}) \text{ or } y^{(n)} \text{or } \frac{d^ny}{dx^n} \text{ or } \frac{d^n}{dx^n}f(x) \text{ or } D^nf(x) \text{ or } D^ny$$

$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$ 이고, f'의 정의역은 위의 극한이 존재하는 모든 x들의 집합

$$f' \; or \; \frac{df}{dx} \;\; or \;\; Df$$

혹은 $\mathrm{y} = \mathrm{f}(x)$ 이면,

$$f'(\mathrm{x}) \text{ or } y' \text{ or } \frac{dy}{dx} \text{ or } Dy$$

ex. Find the derivative of the function $f(x) = x^2 - 8x + 9$ at the number a.

ex. Find an equation of the tangent line to the parabola $y = x^2 - 8x + 9$ at the point $(3, -6)$.

ex. Find the derivative of the function $f(x) = x^2 - 8x + 9$ at the number a

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

$$= \lim_{h \to 0} \frac{\left[(a+h)^2 - 8(a+h) + 9\right] - [a^2 - 8a + 9]}{h}$$

$$= \lim_{h \to 0} \frac{a^2 + 2ah + h^2 - 8a - 8h + 9 - [a^2 - 8a + 9]}{h}$$

$$= \lim_{h \to 0} 2a + h - 8 = 2a - 8$$

ex. Find an equation of the tangent line to the parabola
$y = x^2 - 8x + 9$ at the point $(3, -6)$.

$$f'(a) = 2a - 8 \Rightarrow f'(3) = 6 - 8 = -2$$

$$y - (-6) = (-2)(x - 3)$$

$$y = -2x$$

ex. We found that the first derivative $f'(x) = 2x - 8$ for $y = x^2 - 8x + 9$. Find the second derivative is

$$f''(x) = \frac{df'}{dx}$$

ex. We found that the first derivative $f'(x) = 2x - 8$ for $y = x^2 - 8x + 9$. Find the second derivative is

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

$$= \lim_{h \to 0} \frac{2(x+h) - 8 \ - (2x - 8)}{h}$$

$$= \lim_{h \to 0} \frac{2h}{h}$$

$$= 2$$

Note. If $f(x), g(x)$ are both differentiable,

- Product rule

$$\frac{df}{dx}[f(x)g(x)] = f(x)\,\frac{d}{dx}[g(x)] + g(x)\frac{d}{dx}[f(x)]$$

- Quotient rule

$$\frac{df}{dx}\left[\frac{f(x)}{g(x)}\right] = \frac{\left\{g(x)\frac{d}{dx}[f(x)] - f(x)\,\frac{d}{dx}[g(x)]\right\}}{[g(x)]^2}$$

Note. If $f(x), g(x)$ are both differentiable,

- Chain rule

$$F'(x) = f'\big(g(x)\big) \cdot g'(x)$$

- Leibniz notation, y=$f(u)$, $u = g(x)$,

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$$

Note. If $f(x), g(x)$ are both differentiable,

- Power rule and chain rule

$$\frac{d}{dx}(u^n) = nu^{n-1}\frac{du}{dx}$$

or

$$\frac{d}{dx}(g(x)^n) = n[g(x)]^{n-1} \cdot g'(x)$$

Implicit differentiation

$$y = f(x) \quad or \quad g(x, y) = 0$$

Ex. $x^3 + y^3 = 6xy$, find $\frac{dy}{dx}$. $(y = f(x))$

$$\frac{\mathrm{d}\left(\mathrm{x}^3 + y^3\right)}{dx} = \frac{d(6\,xy)}{dx}$$

Ex. $x^3 + y^3 = 6xy$, find $\frac{dy}{dx}$. $(y = f(x))$

$$\frac{\mathrm{d}(\mathrm{x}^3 + y^3)}{dx} = \frac{d(6\,xy)}{dx} \Rightarrow \frac{\mathrm{d}(\mathrm{x}^3) + d(y^3)}{dx} = \frac{d(6\,xy)}{dx}$$

- $\frac{\mathrm{d}x^3}{dx} = 3x^2$
- $\frac{\mathrm{d}y^3}{dx} = 3y^2\,\frac{dy}{dx}$
- $\frac{d(6\,xy)}{dx} = 6y + 6x\,\frac{dy}{dx}$

$$\therefore\ \frac{dy}{dx} = \frac{2y - x^2}{y^2 - 2x}$$

Definition A function f is one-to-one function if it never takes on the same value twice

$$f(x_1) \neq f(x_2) \quad \text{whenever } x_1 \neq x_2$$

A function f is one-to-one if and only if no horizontal line intersects its graph more than once.

Ex. Is the function $f(x) = x^3$ one-to-one ?

<u>Definition</u> Let f be a one-to-one function with domain A and range B. Then its **inverse function** f^{-1} has domain B and range A and is defined by

$$f^{-1}(y) = x \iff f(x) = y$$

for any y in B

caution. $f^{-1}(\mathrm{x}) \neq \frac{1}{f(x)} (= [f(x)]^{-1})$

Ex. If $f(3) = 7$. Find $f^{-1}(7) =$? (Suppose that it is exists)

Ex. $y = 2x$ find f^{-1}, $(f^{-1})'(2)$.

$$x = 2y \Rightarrow y = \frac{1}{2}x = f^{-1}(x),$$
$$(f^{-1})'(2) \Rightarrow (f^{-1})'(2) = \frac{1}{2},$$

Ex. $f(x) = 2x + \cos x$, find $(f^{-1})'(1)$.

Note. If f is a one-to-one differentiable function with inverse function f^{-1} and $f'(f^{-1}(\mathrm{a})) \neq 0$, then the inverse function is differentiable at a and

$$(f^{-1})'(\mathrm{a}) = \frac{1}{f'(f^{-1}(\mathrm{a}))}$$

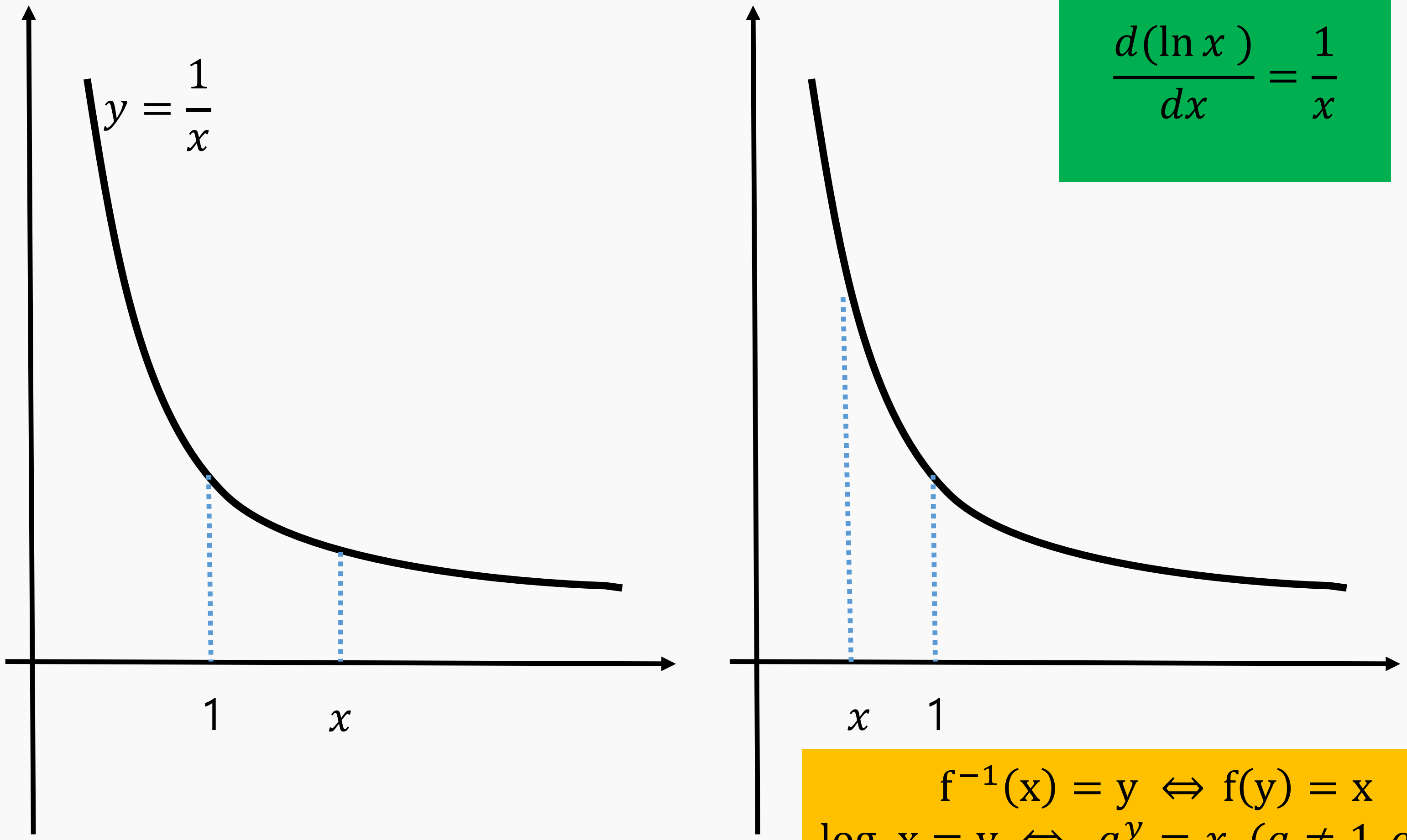
- Leibniz notation, $y = f^{-1}(x), f(x) = x$,

$$\frac{dy}{dx} = \frac{1}{\frac{dx}{dy}}$$

Ex. $f(x) = 2x + \cos x$, find $(f^{-1})'(1)$.

$$Since \quad (f^{-1})'(\mathrm{a}) = \frac{1}{f'(f^{-1}(\mathrm{a}))},$$

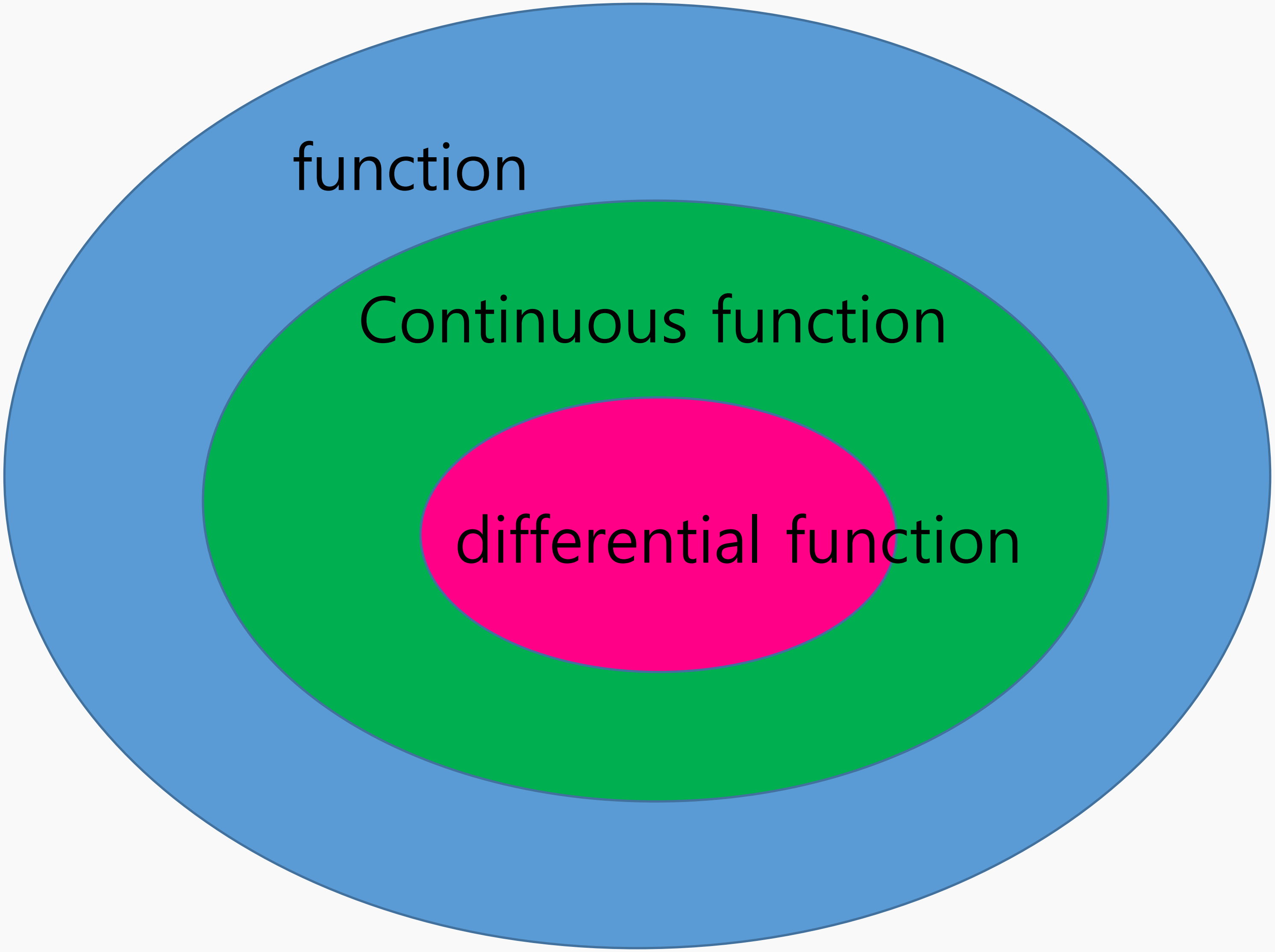
$$(f^{-1})'(1) = \frac{1}{f'(f^{-1}(1))} = \frac{1}{f'(0)} = \frac{1}{2 - \sin x} = \frac{1}{2}$$



$$\frac{d(\ln x)}{dx} = \frac{1}{x}$$

$$f^{-1}(x) = y \iff f(y) = x$$
$$\log_a x = y \iff a^y = x \ (a \neq 1, a > 0)$$
$$\ln x = \log_e x = y \iff e^y = x$$

$\ln x := \ x \geq 1$ 일때, 1과 x 사이의 구간에서 곡선 $y = \frac{1}{x}$ 과 x 축 사이의 면적,
즉, $x > 0$ 에 대해서만 정의, $\ln 1 = 0$.


function
Continuous function
differential function

Indefinite Integral

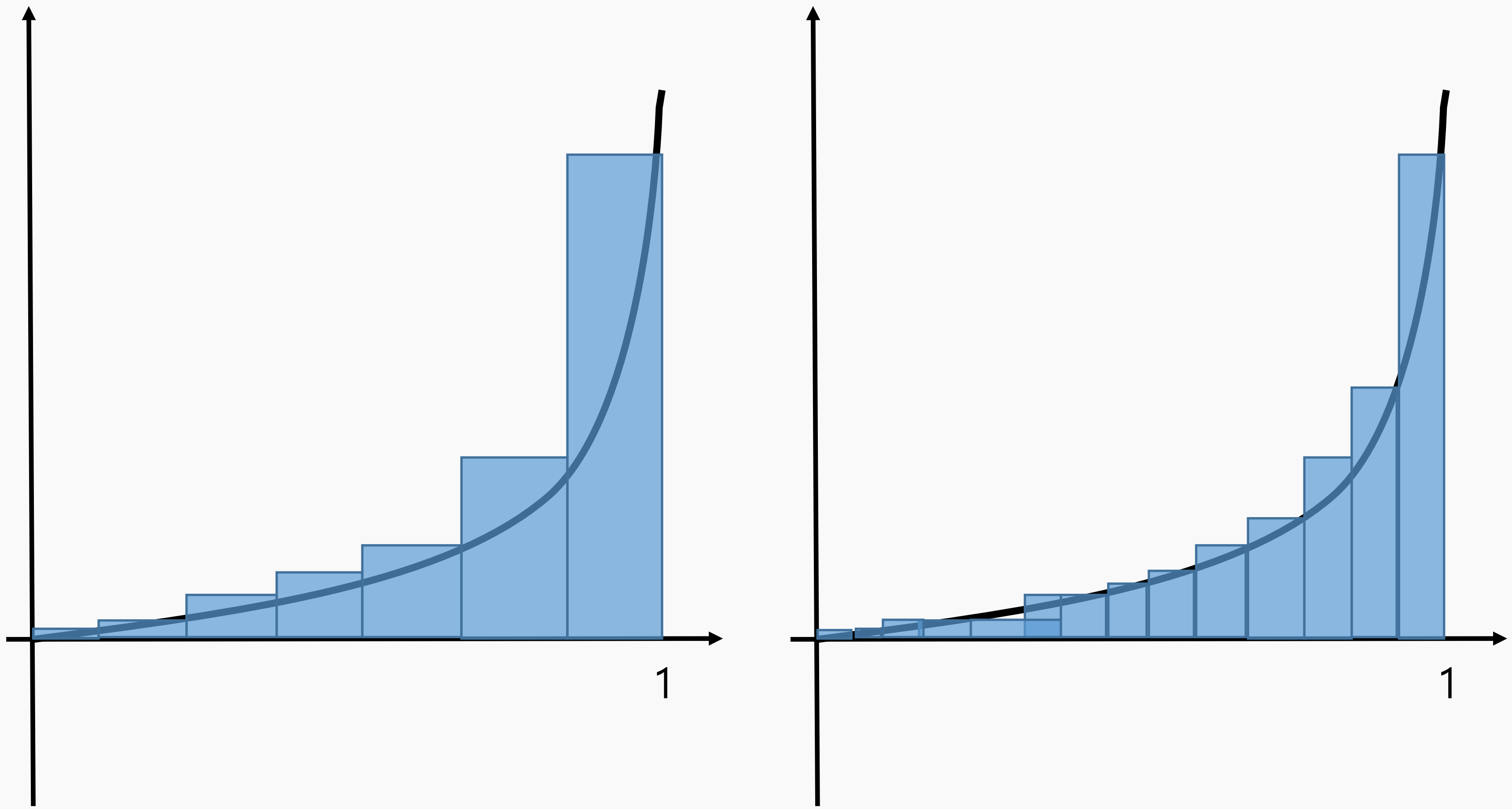
Definition Let F is a **antiderivative** of a function f is a differential function F whose derivative is equal to f.

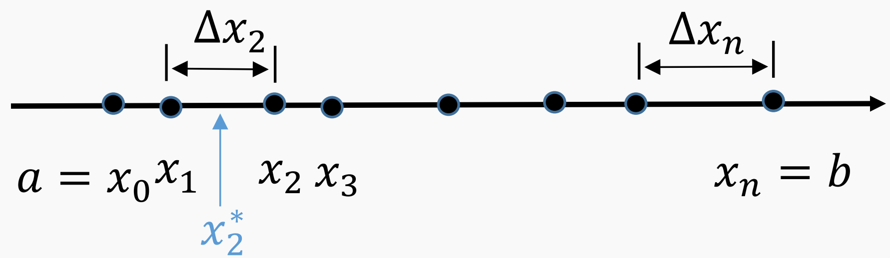
$$F'(x) = f(x)$$

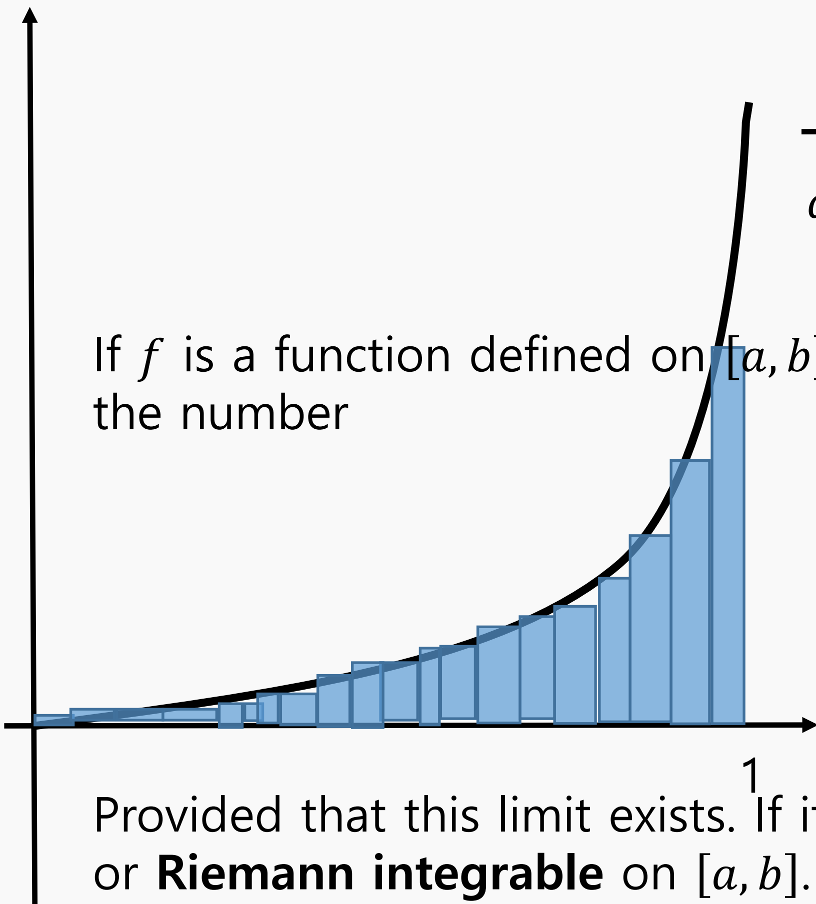
$$(F(x)+1)' = f(x), (F(x)+2)' = f(x), \ (F(x)+c)' = f(x), \ldots$$

$$\Rightarrow \int f(x)dx = F(x) + c$$

Area







If f is a function defined on $[a, b]$, the **definite integral of f** from $a\ to\ b$ is the number

$$\int_a^b f(x)dx = \lim_{\max \Delta x_i \to 0} \sum_{i=1}^{n} f(x_i^*)\Delta \mathrm{x_i}$$

Provided that this limit exists. If it does exist, we say that f is **integrable** or **Riemann integrable** on $[a, b]$.

정적분 또는 리만적분

$$\int_a^b f(x)dx = \int_a^b f(s)ds = \int_a^b f(t)dt = \int_a^b f(u)du$$

$$\int_a^b f(x)dx = \int_a^b f$$

함수 f 가 구간 $[a,b]$에서 유계이고, 유한개의 점을 제외한 모든 점에서 연속일때도 f 는 $[a,b]$에서 적분가능하다.

Partial Derivatives

Multivariate function

<u>Definition</u> A **function f of several variables** is a rule that assigns to each ordered pair of real numbers (x, y) in a set D a unique real number denoted by $f(\mathrm{x}, \mathrm{y})$. The set D is the **domain** of f and its **range** is the set of values that f takes on, that is, $\{f(x, y) | (x, y) \in D\}$

dependent variables

Ex. f : $R \to R$ defined by $\mathrm{y} = f(x)$

f : $R^2 \to R$ defined by $\mathrm{z} = f(x, y)$

independent variables

f : $R^n \to R$ defined by $\mathrm{z} = f(x_1, x_2, \dots, x_n) = f(\boldsymbol{x})$

Review of 1-variable

Definition The limit of $f(x)$,

as x approaches a, equals L

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

$$\Leftrightarrow \forall \epsilon > 0, \exists \delta > 0 \ s.t$$
$$0 < |x - a| < \delta \ \Rightarrow |f(x) - L| < \epsilon$$

Limit of 2-variables

<u>Definition</u> The limit of $f(x, y)$,

as x approaches (a, b), equals L

$$\lim_{(x,y)\to(a,b)} f(x, y) = L$$

$$\Leftrightarrow \forall\epsilon > 0, \exists\delta > 0 \ s.t$$

$$0 < \sqrt{(x-a)^2 + (y-a)^2} < \delta \ \Rightarrow |f(x, y) - L| < \epsilon$$

Limit of n-variables

Definition If f is definded on a subset D of R^n,

then

$$\lim_{\boldsymbol{x} \to \boldsymbol{a}} f(\boldsymbol{x}) = L$$

means that for every $\epsilon > 0$ there is a corresponding number $\delta > 0$ such that

if $\boldsymbol{x} \in D$ and $0 < |\boldsymbol{x} - \boldsymbol{a}| < \delta$ then $|f(\boldsymbol{x}) - L| < \epsilon$

The function is **continuous at** $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ if

$$\lim_{(x,y,z) \to (a,b,c)} f(x, y, z) = f(a, b, c)$$

Definition If f is a function of two variables x and y. Suppose we let only x vary while keeping y fixed, say $y = b$, where b is a constant.

then we are considering a function of a single variable x, $g(x) = f(x, b)$. If g has a derivative at a, then we call it the **partial derivative of f with respect to x at (a, b).**

$$f_x(a, b) = g'(a)$$

By the definition

$$g'(a) = \lim_{h \to 0} \frac{g(a+h) - g(a)}{h}$$

$$f_x(a,b) = \lim_{h \to 0} \frac{f(a+h,b) - f(a,b)}{h}$$

$$f_y(a,b) = \lim_{h \to 0} \frac{f(a,b+h) - f(a,b)}{h}$$

If $\mathrm{f} : R^n \rightarrow R$ is a function of n variables

$$f_{x_i}(x_1, x_2, \cdots, x_n) = \lim_{h \to 0} \frac{f(x_1, \cdots, x_i + h, x_n) - f(x_1, x_2, \cdots, x_n)}{h}$$

$$f_{x_i}(x_1, x_2, \cdots, x_n) = f_i = \frac{\partial f}{\partial x_i} = \frac{\partial}{\partial x_i} f(x_1, x_2, \cdots, x_n) = \frac{\partial z}{\partial x_i} = D_i f$$

Ex. If $f(x, y) = x^3 + x^{2y^3} - 2y^2$, find $f_x(2,1)$.

If $\mathrm{f} : R^n \rightarrow R$ is a function of n variables,

the gradient vector, $\boldsymbol{\nabla f}$

$$\nabla f(\boldsymbol{x}) = \left(\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_n}\right)$$

Check. Product rule, Chain rule!

A **vector-valued function** or **vector function**, is simply a function whose domain is a set of real numbers and whose range is a set of vectors.

$$\boldsymbol{F}: R \rightarrow R^m$$

For every number t in the domain of $\boldsymbol{F}$ there is a unique vector in V_m denoted by $\boldsymbol{F}$(t),

$$\boldsymbol{F}(\mathrm{t}) = (f_1(t), f_2(\mathrm{t}), \cdots, f_m(t))$$

From $\boldsymbol{F}(\mathrm{t}) =< f_1(t)\ , f_2(\mathrm{t}), \cdots, f_m(t) >$

$$\boldsymbol{\nabla F}: R \ \rightarrow \ R^m$$

$$\boldsymbol{\nabla}\, \boldsymbol{F}(\mathrm{t}) = (\, f'_1(t)\ , f'_2(\mathrm{t}), \cdots, f'_m(t))$$

Vector valued multivariate Ft.

From $\boldsymbol{F}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \cdots, f_m(\mathbf{x}))$, $\mathbf{x} \in \boldsymbol{R^n}$

$$\boldsymbol{F}: R^n \to R^m$$

$$\boldsymbol{\nabla} \boldsymbol{F}(\mathbf{x}) = (\boldsymbol{\nabla} f_1(\mathbf{x}), \boldsymbol{\nabla} f_2(\mathbf{x}), \cdots, \boldsymbol{\nabla} f_m(\mathbf{x}))$$

$$= \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial x_n} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix} = \mathrm{d}f_i^T$$

From $\boldsymbol{F}(\mathbf{x}) = (f_1(\mathbf{x}),f_2(\mathbf{x}),\cdots,f_m(\mathbf{x}))$, $\mathbf{x} \in \boldsymbol{R}^{\boldsymbol{n}}$

$$\boldsymbol{F}: R^n \rightarrow R^m$$

Jacobian matrix,

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix} = \boldsymbol{J}_{ij} = \frac{\partial f_i}{\partial x_j}$$

Ex. $\boldsymbol{F}: R^2 \ \rightarrow \ R^2$

$$\boldsymbol{F}\,(x,y) = \begin{bmatrix} x^2 y \\ 5x + \sin y \end{bmatrix} = \begin{bmatrix} f_1(x,y) \\ f_2(x,y) \end{bmatrix}$$

Jacobian matrix, $\boldsymbol{J}_{ij} = \frac{\partial f_i}{\partial x_j} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 2xy & x^2 \\ 5 & \cos y \end{pmatrix}$

From $\boldsymbol{F}(\mathbf{x}) = f(\mathbf{x})$, $\mathbf{x} \in \boldsymbol{R^n}$

$$\boldsymbol{F}: R^n \rightarrow R$$

Hessian matrix,

$$\begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{pmatrix} = \boldsymbol{H_{ij}} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

Ex. $\boldsymbol{F}: R^2 \rightarrow R$

$$\boldsymbol{F}\,(x, y) = x^2 y$$

Hessian matrix,

$$\boldsymbol{H_{ij}} = \frac{\partial^2 f}{\partial x_i \partial x_j} = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} \end{pmatrix} = \begin{pmatrix} 2y & 2x \\ 2x & 0 \end{pmatrix}$$

Q & A