Chapter 5. Vector calculus.

Fonction $\Rightarrow$ $f: R^P \to R$. $\quad$ R^P : codomain

$x \to f(x)$ $\quad$ $f(x)$: image = range = codomain.

° Differentiation of Univariate function.

변수 한개가 input인 함수 (ex. $y = f(x)$)

- Difference quotient : $\dfrac{\delta y}{\delta x} := \dfrac{f(x+\delta x) - f(x)}{\delta x}$ $\quad$ 그래프상에서 2개의 점 가운데 기울기.

  ↳ Derivative : $\dfrac{df}{dx} := \lim_{h \to 0} \dfrac{f(x+h) - f(x)}{h}$ $\quad$ x에서 미분.

- Taylor polynomial

$$T_n(x) := \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k$$

x_0에서 f함수를 K번 미분.

$\Rightarrow$ an approximation of a function.

↳ Taylor series. (Taylor polynomial에서 n을 ∞로 보냄)

$$T_\infty(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k$$

$\Rightarrow$ $x_0 = 0$: Maclaurin series

$f(x) = T_\infty(x)$: analytic.

- Differentiation rules.

① $(f(x)g(x))' = f'(x)g(x) + f(x)g'(x)$

② $\left(\dfrac{f(x)}{g(x)}\right)' = \dfrac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}$.

③ $(f(x) + g(x))' = f'(x) + g'(x)$

④ $(g(f(x)))' = g'(f(x))f'(x)$.

○ Partial Differentiation and Gradient.

$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x \in R^n$, $\quad f: R^n \to R$ $\quad f(x_1, x_2 \cdots x_n)$

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

$\vdots$

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

$$\Rightarrow \nabla_x f = \text{grad } f = \frac{df}{dx} = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots \frac{\partial f}{\partial x_n}\right] \in R^{1 \times n}.$$

- Basic rule of partial diffentiation.

① $\frac{\partial}{\partial x}\left(f(x)g(x)\right) = \frac{\partial f}{\partial x} g(x) + f(x)\frac{\partial g}{\partial x}$

② $\frac{\partial}{\partial x}\left(f(x)+g(x)\right) = \frac{\partial f}{\partial x} + \frac{\partial g}{\partial x}$

③ chain rule : $\frac{\partial}{\partial x}(g \circ f)(x) = \frac{\partial g}{\partial f} \cdot \frac{\partial f}{\partial x}$

ex) x_1, x_2 는 t에 대한 함수.

$$\frac{df}{dt} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial t} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial t}.$$

ex) x_1, x_2: s와 t에 대한 함수.

$$\frac{d(f)}{d(s,t)} = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}\right]\begin{bmatrix}\frac{\partial x_1}{\partial s} & \frac{\partial x_1}{\partial t} \\ \frac{\partial x_2}{\partial s} & \frac{\partial x_2}{\partial t}\end{bmatrix}$$

∘ Gradient of Vector-Valued Function.

- Vector-Valued Function : $f: \mathbb{R}^n \to \mathbb{R}^m$.

- Jacobian

$$J = \nabla_x f = \frac{df(x)}{dx} = \begin{bmatrix}\frac{\partial f_1(x)}{\partial x_1} & & \\ \vdots & & \\ \frac{\partial f_m(x)}{\partial x_1} & \cdots & \frac{\partial f_m(x)}{\partial x_n}\end{bmatrix}_{m\times n.}$$

$x \in \mathbb{R}^n$

$f(x) \in \mathbb{R}^m$

$\underset{\mathbb{R}^m}{f(x)} = \underset{m\times n}{J}\,\underset{n\times 1}{x}.$

∘ Gradient of matrices.

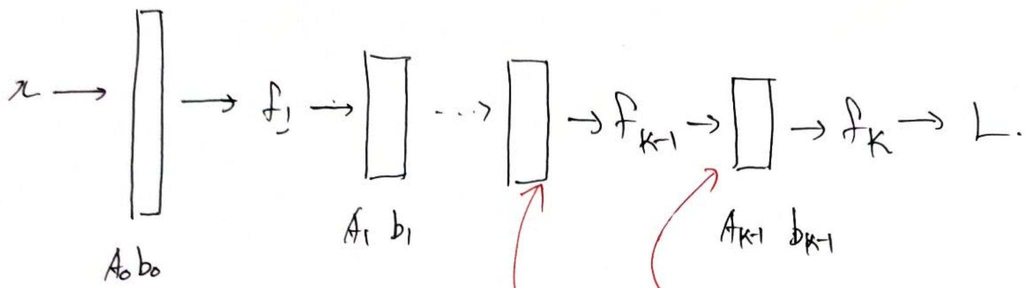
① $A_{m\times n}, B_{p\times q} \to J_{(m\times n)\times(p\times q)}$ $\left(J_{ijkl} = \frac{\partial A_{ij}}{\partial B_{kl}}\right)$

② $A_{m\times n} \xrightarrow{\text{reshape}} A_{mn\times 1}$

$B_{p\times q} \to A_{pq\times 1}$ $\Big\} \to J_{mn\times pq} \xrightarrow{\text{reshape}}$ original shape.

◦ Backpropagation and Automatic Differentiation.

– Backpropagation algorithm ⇒ error function을 gradient하는데 계산을 효율적으로 해줌.



$\theta = \{A_0, b_0, A_1 \cdots b_{K-1}\}$

$\theta_j = \{A_j\ b_j\}$

ex) $f_2 = \sigma_2(A_1 f_1 + b_1)$

$L = \| y - f_K(\theta - x) \|^2$

$$\frac{\partial L}{\partial \theta_{K-1}} = \frac{\partial L}{\partial f_K} \frac{\partial f_K}{\partial \theta_{K-1}}$$

$$\frac{\partial L}{\partial \theta_{K-2}} = \frac{\partial L}{\partial f_K} \cdot \frac{\partial f_K}{\partial f_{K-1}} \cdot \frac{\partial f_{K-1}}{\partial \theta_{K-1}}$$

⋮

$$\frac{\partial L}{\partial \theta_i} = \frac{\partial L}{\partial f_K} \frac{\partial f_K}{\partial f_{K-1}} \cdots \frac{\partial f_{i+2}}{\partial f_{i+1}} \frac{\partial f_{i+1}}{\partial \theta_i}$$

* chain rule을 사용한 방법.
reuse를 많이 함.

◦ Higher-order Derivatives.

Hessian → 모든 2차 derivatives의 모음.

◦ Linearization and Multivariate Taylor Series.

$$f(x) \approx f(x_0) + \underbrace{(\nabla_x f)(x_0)}_{\text{기울기}} (x - x_0)$$

⇐ x_0에서의 linear approximation

Multivariate Taylor series.

$\delta := x - x_0$

$$f(x) = \sum_{k=0}^{\infty} \frac{D_x^k f(x_0)}{k!} \delta^k.$$

approximation에서 2차수 이상을 다룸.
gradient가 아닌 Taylor 방식 적용.

Taylor Polynomial

$$T_n(x) = \sum_{k=0}^{n} \frac{D_x^k f(x_0)}{k!} \delta^k.$$