Lecture 7. Vector Calculus : Differentiation or Integration ⇒확장시 Multi Variate Calculus.

1차 미분, Chain Rule

Difference Quotient ⟶ Partial Derivatives → Optimization, Regression, Dimensionality Reduction, Density Estimation, Classification

Partial Derivatives ⟶ Jacobian, Hessian

2차이상 미분

Taylor Series ← Jacobian, Hessian → Probability.

Function

$f : \mathbb{R}^D \to \mathbb{R}$ $\mathbb{R}^D$: domain

$x \mapsto f(x)$ $\{f(x) : x \in \mathbb{R}^D\}$: image (= codomain)

Differentiation of univariate function

Univariate function: $y = f(x)$, $x, y \in \mathbb{R}$ ⇒독립변수가 1개인 함수.

Difference Quotient 기울기.

$$\frac{\delta y}{\delta x} := \frac{f(x+\delta x) - f(x)}{\delta x}$$ 두점의 기울기 ⇒ Derivative $\frac{df}{dx} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$.

Taylor Series

function f 를 infinite sum of term ⇒ 어떤 point x_0 에 대해 결정되는 f 의 derivative.

$n \to \infty$

Taylor Polynomial : $x = x_0$ 에서 n차 $f : \mathbb{R} \to \mathbb{R}$ 인 함수.

$$T_n(x) := \sum_{k=0}^{n} \underbrace{\frac{f^{(k)}(x_0)}{k!}}_{\text{Coefficient of polynomial}} (x - x_0)^k$$

$\approx f(x)$ but, $f(x)$ 가 n차이하의 polynomial 일때, $T_n(x)$와 같음.

$f(x) = T_\infty(x)$: Analytic

$x_0 = 0$ 인 $T_n(x)$: Maclurin Series

$\frac{f^{(k)}(x_0)}{k!} = a_k$ 일 때 (일반화): Power Series

Partial Derivatives and Gradient.

$x \in \mathbb{R}^n$ $n \times 1$ Vector.

$f: \mathbb{R}^n \to \mathbb{R}$ $x \mapsto f(x)$, $x = [x_1 \cdots x_n]^T$ 의 partial derivative

$$\frac{\partial f}{\partial x_1} = \lim_{h \to 0} \frac{f(x_1 + h, x_2, x_3 \cdots x_n) - f(x)}{h}$$

$\vdots$

$\frac{\partial f}{\partial x_n}$

n개의 derivatives

{ Gradient of f ⇒ $\nabla_x f = \frac{df}{dx} = \left[\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \cdots, \frac{\partial f(x)}{\partial x_n}\right]$ row vector로 표현!
Jacobian

⇓

Chain Rule 에 의한 확장.

$$\frac{\partial f}{\partial s} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial s} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial s}$$

$$\frac{\partial f}{\partial t} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial t} + \frac{\partial f}{\partial x_2}\cdot\frac{\partial x_2}{\partial s}$$

$$\Rightarrow \frac{\partial f}{\partial (s,t)} = \frac{\partial f}{\partial x}\cdot\frac{\partial x}{\partial (s,t)} = \underbrace{\begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{bmatrix}}_{\frac{\partial f}{\partial x}} \underbrace{\begin{bmatrix} \frac{\partial x_1}{\partial s} & \frac{\partial x_1}{\partial t} \\ \frac{\partial x_2}{\partial s} & \frac{\partial x_2}{\partial t} \end{bmatrix}}_{\frac{\partial f}{\partial (s,t)}}$$

(s,t)가 column vector여도 derivative에 의해 row vector로 등일 축 변수에 의해 펼쳐짐

Vector - valued Function.

$f: \mathbb{R}^n \to \mathbb{R}^m$

$x = [x_1 \cdots x_n]^T \in \mathbb{R}^n$

$x \mapsto f(x) = [f_1, f_2, \cdots f_m]^T \in \mathbb{R}^m$

Partial derivative of a vector - valued function.

종속변수에 대해 column vector.

$$J = \nabla_x f = \frac{df(x)}{dx} = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \cdots & \frac{\partial f_1(x)}{\partial x_n} \\ \frac{\partial f_2(x)}{\partial x_1} & & \\ \vdots & & \\ \frac{\partial f_m(x)}{\partial x_1} & & \end{bmatrix}$$

→ 독립변수에 대해 row vector

transformation of variable 에서 많이 사용됨.

$|\det(J)|$ = scaling factor of J.

Gradient of matrices

gradient of $m \times n$ matrix A with respect to (~에 대해) $p \times q$ matrix B.

$\Downarrow$

Result of Jacobian: $m \times n \times p \times q$ 4 dimensional Tensor of J.

연산 방법.

i) B의 element에 대해 partial derivative 진행. $\Rightarrow$ Collate.

ii) reshape(flatten) $A \in R^{m \times n}$ into $A' \in R^{mn}$ 이후, B에 대한 partial derivative 후 원래대로 reshape.

주로 사용처: Back Propagation Algorithm.

↳ loss function에서 DNN에서의 여러 층의 derivative를 구하기 위해 주로 사용.

Higher Order Derivative.

Hessian (2nd order derivative)

↳ twice differentiable function $f(x,y)$

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} := \nabla^2_{x,y} f(x,y)$$

$f: R^n \to R$: Hessian은 $n \times n$ matrix

$f: R^n \to R^m$: Hessian $m \times n \times n$ tensor

Linear Approximation of a function at a point x_0.

$f(x) \simeq f(x_0) + (\nabla_x f)(x_0)(x - x_0)$: x_0 에서는 매우 유사함.

x_0에서 멀면 worse.

Multivariate Taylor series

$f: R^D \to R$

$x \mapsto f(x)$, $x \in R^D$

$f(x)$가 x_0에서 smooth하다고 할 때, difference vector δ

x와 x_0가 $\in R^n$ 인 vector (univariate는 scalar)

$$f(x) = \sum_{k=0}^{\infty} \frac{D_x^k f(x_0)}{k!} \delta^k \quad \text{when } \delta := x - x_0$$

$\Downarrow$

Taylor Polynomial

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