1. Definition of gradient and Jacobian.

① univariate scalar function : $\mathbb{R} \to \mathbb{R}$ 우리가 아는거

(변수 / 결과)

$f(x) = ax^2 + bx \longrightarrow \nabla f(x) = 2ax + b$

* scalar function 미분 - gradient → 세로
  vector function 미분 - Jacobian → 가로

(F : 벡터함수 . f : 스칼라 함수. x : 스칼라 . $\mathbf{x}$: 벡터)

② univariate vector function : $\mathbb{R} \to \mathbb{R}^n$ 벡터함수

$$F(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_n(x) \end{pmatrix}$$

↑ univariable scalar function

$$\nabla F(x) = (\nabla f_1(x), \cdots, \nabla f_n(x))$$

ex) $F(x) = \begin{pmatrix} 3x+1 \\ 2x^2-1 \end{pmatrix} \to \nabla F(x) = (3, 4x)$

③ multivariate scalar function $\mathbb{R}^n \to \mathbb{R}$ 다변수 함수

$$f(\mathbf{x}) = f(x_1, \cdots, x_n)$$

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{pmatrix}$$

편미분 모음

ex) $f(\mathbf{x}) = 3x_1^2 + 4x_2^2 + 5x_1x_2 \to \nabla f(\mathbf{x}) = \begin{pmatrix} 6x_1 + 5x_2 \\ 8x_2 + 5x_1 \end{pmatrix}$

④ multivariate vector function $\mathbb{R}^n \to \mathbb{R}^m$

$$F(\mathbf{x}) = \begin{pmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{pmatrix}$$

$x_1, \cdots, x_n$

↑ multivariate scalar function

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

ex) $F(x) = \begin{pmatrix} 2x_1x_2 + x_2^2 \\ x_1^2 + 2x_1x_2 \end{pmatrix}$

$$\to \nabla F(x) = \begin{pmatrix} 2x_2 & 2x_1 + 2x_2 \\ 2x_1 + 2x_2 & 2x_1 \end{pmatrix}$$

↓ $\nabla f_1(x)$ ↓ $\nabla f_2(x)$

* 미분의 일반적 형태

① $f(x) = ax \longrightarrow \nabla f(x) = a^T$ 해도 무방

② $F(x) = xa \longrightarrow \nabla F(x) = a^T = (a_1, \cdots, a_n)$

스칼라를 뒤로 보내야 함!

$\begin{pmatrix} a_1x \\ \vdots \\ a_mx \end{pmatrix}$ $a = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix}$

$F(x) = g(x)a \longrightarrow \nabla F(x) = \nabla g(x) a^T$

$\begin{pmatrix} g(x)a_1 \\ \vdots \\ g(x)a_m \end{pmatrix}$ $(g'(x)a_1, \cdots, g'(x)a_n)$

③ $f(\mathbf{x}) = Ax = xA \longrightarrow \nabla f(x) = A^T$

$(A = a^T)$ $\begin{pmatrix} a_1^T \\ \vdots \\ a_n^T \end{pmatrix}$

④ $F(x) = Ax = \begin{pmatrix} a_1^Tx \\ \vdots \\ a_m^Tx \end{pmatrix} \longrightarrow \nabla F(x) = A^T$

$(A = a^T)$ $\begin{pmatrix} a_1^T \\ \vdots \\ a_m^T \end{pmatrix}$ $(a_1, \cdots, a_m)$

↑ Linear combination

$\boxed{\therefore \text{함수 } ax \text{ 미분} = a^T}$

2. Gradient of products and ratios (곱의 미분, 몫의 미분)

① $f(x) = \underbrace{g(x)\,h(x)}_{\text{스칼라 함수}} \longrightarrow \nabla f(x) = \nabla g(x)\,h(x) + \nabla h(x)\,g(x)$ 기본! $\mathbb{R} \to \mathbb{R}$

② $\to F(x) = G(x)\,h(x)$ $\mathbb{R} \to \mathbb{R}^m$ (벡터함수 × 스칼라 함수)

$\left(G(x) = \begin{pmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{pmatrix} \right) \longrightarrow F(x) = \begin{pmatrix} g_1(x)\cdot h(x) \\ \vdots \\ g_m(x)\,h(x) \end{pmatrix} \longrightarrow \nabla F(x) = \nabla G(x)\,h(x) + \nabla h(x) \cdot \nabla G(x)^T$ $\mathbb{R} \to \mathbb{R}^{1\times m}$

$\nabla F(x) = (\nabla(g_1(x)h(x)), \cdots, \nabla(g_m(x)h(x)))$

$= (\nabla g_1(x)h(x) + \nabla h(x)g_1(x), \cdots, \nabla g_m(x)h(x) + \nabla h(x)\,g_m(x))$

$= (\nabla g_1(x), \cdots, \nabla g_m(x))h(x) + \nabla h(x)\,(g_1(x), \cdots, g_m(x))$

$\to F(x) = \dfrac{G(x)}{h(x)}$ $\mathbb{R} \to \mathbb{R}^m \longrightarrow \nabla F(x) = \dfrac{\{\nabla G(x)h(x) - \nabla h(x)G(x)^T\}}{(h(x))^2}$ $= \nabla G(x)\,h(x) + \nabla h(x)\,G(x)^T$ $\mathbb{R} \to \mathbb{R}^{1\times m}$

in multivariate function (스칼라×스칼라)

③ $\to f(x) = g(x)\,h(x)$ $\mathbb{R}^n \to \mathbb{R} \longrightarrow \nabla f(x) = \nabla g(x)\,h(x) + \nabla h(x)\,g(x)^T$ $\mathbb{R}^n \to \mathbb{R}^n$

$\nabla f(x) = \begin{pmatrix} \frac{\partial}{\partial x_1}(g(x)h(x)) \\ \vdots \\ \frac{\partial}{\partial x_n}(g(x)h(x)) \end{pmatrix} = \begin{pmatrix} (\frac{\partial}{\partial x_1}g(x))h(x) + (\frac{\partial}{\partial x_1}h(x))g(x) \\ \vdots \\ \end{pmatrix}$

$\to f(x) = \dfrac{g(x)}{h(x)}$ $\mathbb{R}^n \to \mathbb{R} \longrightarrow \nabla f(x) = \dfrac{\nabla g(x)h(x) - \nabla h(x)g(x)^T}{(h(x))^2}$ $\mathbb{R}^n \to \mathbb{R}^n$

④ $F(x) = G(x)\,h(x)$ $\mathbb{R}^n \to \mathbb{R}^m$ (벡터함수 × 스칼라 함수)

$x_1 \cdots x_n$

$\to \nabla F(x) = \underbrace{\nabla G(x)}_{n\times m}\,\underbrace{h(x)}_{\substack{\text{스칼라 함수} \\ \cancel{m\times}\ 1\times 1}} + \underbrace{\nabla h(x)}_{n\times 1}\,\underbrace{G(x)^T}_{1\times m}$ $\mathbb{R}^n \to \mathbb{R}^{n\times m}$ size 체크!

$\to F(x) = \dfrac{G(x)}{h(x)}$ $\mathbb{R}^n \longrightarrow \mathbb{R}^m \longrightarrow \nabla F(x) = \dfrac{\nabla G(x)h(x) - \nabla h(x)\,G(x)^T}{(h(x))^2}$ $\mathbb{R}^n \to \mathbb{R}^{n\times m}$

⑤ 내적 (Inner product)

$f(x) = G(x)^T H(x)$ $\mathbb{R}^n \to \mathbb{R} \to \nabla f(x) = \nabla G(x)\,H(x) + \nabla H(x)\,G(x)$

~~내적만 T없음! 미분에서 사라짐~~

$\boxed{\text{T없이 생각하고 뒤에도 T 없음}}$

* 일반적 형태

② $F(x) = (ax)\,h(x) \to \nabla F(x) = a^T h(x) + \nabla h(x)\,(ax)^T$

$F(x) = a(\underbrace{x\,h(x)}_{\text{스칼라}}) \longrightarrow \nabla F(x) = \{h(x) + \nabla h(x)\,x\}\,a^T$

벡터를 앞으로!

③ $f(x) = (a^T x)(c^T x) \longrightarrow \nabla f(x) = a(c^T x) + c(a^T x)$

④ $F(x) = Ax\,(a^T x) \longrightarrow \nabla F(x) = A^T(a^T x) + a(Ax)^T$

$F(x) = c\,(a^T x) \longrightarrow \nabla F(x) = 0\,(\cancel{a^T x}) + ac^T = ac$

$F(x) = (ca^T)x \longrightarrow \nabla F(x) = (ca^T)^T$

⑤

$f(x) = (Ax)^T(Bx)$

$\to \nabla f(x) = A^T(Bx) + B^T(Ax)$

$f(x) = x^T Cx = (C^T x)^T I x$ 일부러 T 취해서 내적으로 만듦

$\to \nabla f(x) = C(Ix) + I(C^T x)$

$\boxed{\begin{array}{l} \therefore \nabla f(x) = \nabla g(x)\,h(x) + \nabla h(x)\,g(x)^T. \\ \text{뭐끼리 곱인지 (벡터인지 스칼라인지 잘 확인 후 공식 적용)} \end{array}}$

$f: a^T x = x^T a$

$\nabla f: a$ 동일

(x^T 되있으면 그대로 a)

추가 ex)

1. $f(x) = x^T a c^T x$, $x \in \mathbb{R}^n$. $a, c \in \mathbb{R}^n$: $\mathbb{R}^n \to \mathbb{R}$

① 변수, 결과 확인
② 스칼라 벡터곱 : 벡터가 무조건 앞에
③ size check

$= (ax^T)(c^T x)$ x뒤가 미분 편함, 결과 스칼라니 바꿔도 가능

$\nabla f(x) = a(c^T x) + c\underbrace{(ax^T)^T}_{\text{스칼라}}$

$= (ac^T + ca^T)x$

2. $F(x) = \underbrace{x^T a}_{\text{스칼라}} Ax$ → 뒤로

$= (Ax)(x^T a)$

$\nabla F(x) = A^T(x^T a) + a(Ax)^T$

3. $f(x) = x^T Ax$ 벡터

$= x^T A I x$

$= (Ax)^T I x$

$\nabla f(x) = Ax + IA^T x$ ~

3. Chain rules (연쇄법칙)

① $f(x) = g(h(x))$. $x \in \mathbb{R}$. ~~스칼라~~ $\mathbb{R} \to \mathbb{R}$ $\to \nabla f(x) = \nabla g(h(x)) * \nabla h(x) = \nabla h(x) \cdot \nabla g(h(x))$

② $F(x) = G(h(x))$ $x \in \mathbb{R}$ $\mathbb{R} \to \mathbb{R} \to \mathbb{R}^m$

$\to \nabla F(x) = \nabla h(x) \nabla G(h(x))$ $\mathbb{R} \to \mathbb{R}^{1\times m}$

③ $f(x) = g(h(x))$. $x \in \mathbb{R}^n$ $\mathbb{R}^n \to \mathbb{R} \to \mathbb{R}$

$\to \nabla f(x) = \nabla h(x) \nabla g(h(x))$ $\mathbb{R}^n \to \mathbb{R}$

④ $F(x) = G(h(x))$. $x \in \mathbb{R}^n$ $\mathbb{R}^n \to \mathbb{R} \to \mathbb{R}^m$

$\to \nabla F(x) = \nabla h(x) \nabla G(h(x))$ $\mathbb{R}^n \to \mathbb{R} \to \mathbb{R}^{n\times m}$

* 일반적 형태

벡터함수는
전체 미분 개념이 아니라 앞에서 미분하고 뒤에서 미분하는 느낌 : a가 앞에 있을 때!

② $F(x) = a(x^2+x) \to \nabla F(x) = (2x+1)a^T$
$F(x) = ah(x) \to \nabla F(x) = \nabla h(x) \cdot a^T$

③ $f(x) = (a^T x)^2 \Longrightarrow \nabla f(x) = a \cdot 2(a^T x)$
$f(x) = g(a^T x) \to \nabla f(x) = a \nabla g(a^T x)$
이게 내가 아는거

④ $F(x) = a(c^T x) \to \nabla F(x) = ca^T$
$F(x) = (ac^T)x \to \nabla F(x) = ca^T$
$F(x) = a(c^T x) \longrightarrow \nabla F(x) = 0(c^T x) + ca^T$

곱의 미분이랑 결과 동일!

4. statistical Example.

① Linear regression

$$L(\beta) = \sum_{i=1}^{n} (y_i - x_i^T\beta)^2 = \|y - X\beta\|^2$$

$$= y^Ty - 2y^TX\beta + \beta^TX^TX\beta$$

$$= y^Ty - 2y^TX\beta + (X\beta)^T(X\beta)$$

$$\rightarrow \nabla L(\beta) = -2X^Ty + 2X^TX\beta = 0 \quad \left(= \sum_{i=1}^{n} \left(-2x_i(y_i - x_i^T\beta)\right) \right)$$

$$\left(\nabla^2 L(\beta) = \sum_{i=1}^{n} 2x_ix_i^T = \underline{2X^TX} \right)$$

real symmetric

② Logistic regression

$$L(\beta) = \sum_{i=1}^{n} \left\{ -y_i x_i^T\beta + \log(1 + e^{x_i^T\beta}) \right\}$$

$$\rightarrow \nabla L(\beta) = \sum_{i=1}^{n} \left\{ -(y_i x_i^T)^T + x_i e^{x_i^T\beta} \times \frac{1}{1 + e^{x_i^T\beta}} \right\}$$

$$= \sum_{i=1}^{n} \left(-y_i + \frac{e^{x_i^T\beta}}{1 + e^{x_i^T\beta}} \right) x_i$$

$$\left(\nabla^2 L(\beta) = \sum_{i=1}^{n} e^{x_i^T\beta} \times \left(\frac{1}{1 + e^{x_i^T\beta}} \right)^2 x_i x_i^T \right.$$