선형대수학 정리 (for A.I)

연쇄법칙 (Chain Rule)

연쇄법칙은 합성함수(Composition function)의 도함수를 구할 때, 유용하게 사용된다.

합성함수란 두 개 이상의 함수를 연결하여, 하나의 함수로 만드는 연산을 말한다.

연쇄법칙은 Backpropagation의 알고리즘의 근간을 이룬다.

ex) vector $\vec{y}=[y_1\ y_2]^T$를 변수로 하는 다변수 Scalar 함수 $f(\vec{y})$의 미분 df는 아래와 같이 계산된다.

$$df=\frac{\partial f}{\partial y_1}dy_1+\frac{\partial f}{\partial y_2}dy_2 \quad \cdots (1)$$

ex2) 여기에 더하여 y_1, y_2도 $\vec{x}=[x_1, x_2, x_3]^T$의 함수일 때, $y_1(\vec{x}), y_2(\vec{x})$의 미분은 아래와 같다.

$$\begin{aligned} dy_1 &= \frac{\partial y_1}{\partial x_1}dx_1+\frac{\partial y_1}{\partial x_2}dx_2+\frac{\partial y_1}{\partial x_3}dx_3 \\ dy_2 &= \frac{\partial y_2}{\partial x_1}dx_1+\frac{\partial y_2}{\partial x_2}dx_2+\frac{\partial y_2}{\partial x_3}dx_3 \end{aligned} \quad \cdots (2)$$

$$df=\frac{\partial f}{\partial y_1}\left(\frac{\partial y_1}{\partial x_1}dx_1+\frac{\partial y_1}{\partial x_2}dx_2+\frac{\partial y_1}{\partial x_3}dx_3\right)+\frac{\partial f}{\partial y_2}\left(\frac{\partial y_2}{\partial x_1}dx_1+\frac{\partial y_2}{\partial x_2}dx_2+\frac{\partial y_2}{\partial x_3}dx_3\right) \quad \cdots (3)$$

결국 f는 $\vec{x}$의 함수이기에 아래와 같이 표현할 수 있다.

$$df=\frac{\partial f}{\partial x_1}dx_1+\frac{\partial f}{\partial x_2}dx_2+\frac{\partial f}{\partial x_3}dx_3 \quad \cdots (4)$$

식(4)과 식(3)이 같다는 것에 아래의 관계가 정립된다.

$$\begin{aligned} \frac{\partial f}{\partial x_1} &= \frac{\partial f}{\partial y_1}\cdot\frac{\partial y_1}{\partial x_1}+\frac{\partial f}{\partial y_2}\cdot\frac{\partial y_2}{\partial x_1} \\ \frac{\partial f}{\partial x_2} &= \frac{\partial f}{\partial y_1}\cdot\frac{\partial y_1}{\partial x_2}+\frac{\partial f}{\partial y_2}\cdot\frac{\partial y_2}{\partial x_2} \\ \frac{\partial f}{\partial x_3} &= \frac{\partial f}{\partial y_1}\cdot\frac{\partial y_1}{\partial x_3}+\frac{\partial f}{\partial y_2}\cdot\frac{\partial y_2}{\partial x_3} \end{aligned} \quad \cdots (5)$$

식(5)를 vector & matrix를 사용하여 표현하면 아래와 같다.

$$\begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \frac{\partial f}{\partial x_3} \end{bmatrix} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} \\ \frac{\partial y_1}{\partial x_3} & \frac{\partial y_2}{\partial x_3} \end{bmatrix} \begin{bmatrix} \frac{\partial f}{\partial y_1} \\ \frac{\partial f}{\partial y_2} \end{bmatrix} = \frac{df}{d\vec{x}} = \underbrace{\boxed{\frac{d\vec{y}}{d\vec{x}}}}_{\text{Jacobian}} \cdot \underbrace{\boxed{\frac{df}{d\vec{y}}}}_{\text{Gradient}}$$

※ Gradient and Jacobian.

Gradient

어떤 다변량 함수 $f(x_1, x_2, x_3, \cdots, x_n)$이 있을 때, f의 gradient는 다음과 같이 정의된다.

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n}\right)$$

gradient는 위의 식과 같이 각 변수로의 일차 편미분 값으로 구성되는 Vector이다.

이 Vector는 f의 값이 가장 가파르게 증가하는 방향을 나타낸다.

'Vector의 크기'는 그 증가의 가파른 정도(기울기)를 나타낸다.

ex) $f(x,y) = x^2 + y^2$의 gradient를 구하면

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) = (2x, 2y)$$

Gradient에 (-)를 취하면, $-\nabla f$는 f값이 가장 가파르게 감소하는 방향을 나타낸다.

tip. 영상처리에서 gradient는 영상의 edge 및 edge의 방향을 찾는 용도로 활용될 수 있다.

이미지 $I(x,y)$를 (x,y)에서의 pixel의 밝기를 나타내는 함수로 보고

각 pixel 위치에서 gradient의 크기와 방향을 구하면, 해당 pixel이 얼마나 edge에 가까운지, 그리고 edge의 방향이 어떤지 쉽게 알 수 있다.

Jacobian.

어떤 $F: \mathbb{R}^n \to \mathbb{R}^m$ 함수 $F(x_1, x_2, x_3, \cdots, x_n) = (f_1(x_1, x_2, \cdots, x_n), \cdots, f_m(x_1, x_2, \cdots, x_n))$

Jacobian 행렬은 다음과 같다.

$$J_F = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \quad m \times n \text{ matrix.}$$

Jacobian은 어떤 다변수 vector 함수(Vector-valued function of multiple variables)에 대한 일차미분으로 볼 수 있다.

앞서 나온 gradient나 jacobian 모두 함수에 대한 일차미분을 나타낸다는 점에서 동일하다.

다만 gradient 다변수 Scalar 함수에 대한 일차미분인 반면, Jacobian은 다변수 Vector 함수에 대한 일차미분이다.

위의 식(7)을 다변수 합성함수의 chain Rule (연쇄법칙) 이라 한다.

$$x \xrightarrow{g} y \xrightarrow{f} z \qquad \frac{dz}{d\vec{x}} = \frac{d\vec{y}}{d\vec{x}}\frac{dz}{d\vec{y}}$$

연쇄 법칙은 복합함수의 구성함수가 2개 이상일 때도 적용할 수 있다.

ex) $\vec{x} \longrightarrow \vec{y} \longrightarrow \vec{v} \longrightarrow \cdots\cdots \longrightarrow \vec{w} \longrightarrow z$

$$\frac{dz}{d\vec{x}} = \underbrace{\boxed{\frac{d\vec{y}}{d\vec{x}} \cdot \frac{d\vec{v}}{d\vec{y}} \cdots}}_{\text{Jacobian}} \underbrace{\boxed{\frac{dz}{d\vec{w}}}}_{\text{gradient}}$$

Scalar 함수를 행렬로 미분하는 방법.

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \in \mathbb{R}^{m\times n}$$

행렬이 주어지고 scalar 함수 $f(X)$가 주어졌을 때, $f(X)$에 대한 X의 미분은 다음과 같다.

$$\frac{df}{dX} = \begin{bmatrix} \frac{\partial f}{\partial x_{11}} & \frac{\partial f}{\partial x_{12}} & \cdots & \frac{\partial f}{\partial x_{1n}} \\ \frac{\partial f}{\partial x_{21}} & \frac{\partial f}{\partial x_{22}} & \cdots & \frac{\partial f}{\partial x_{2n}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f}{\partial x_{m1}} & \frac{\partial f}{\partial x_{m2}} & \cdots & \frac{\partial f}{\partial x_{mn}} \end{bmatrix} \in \mathbb{R}^{m\times n}$$

$$\frac{df}{dX} = \begin{bmatrix} w_{11} & w_{21} & w_{31} & \cdots & w_{m1} \\ w_{12} & w_{22} & w_{32} & \cdots & w_{m2} \\ \vdots & \vdots & \vdots & & \vdots \\ w_{1n} & w_{2n} & w_{3n} & \cdots & w_{mn} \end{bmatrix} = W^T$$

Vector 함수를 행렬로 미분하는 방법.

Scalar 함수를 vector로 미분하면, 그 결과가 vector로 나온다.

Vector를 vector로 미분하면, 그 결과가 matrix(Jacobian)으로 나온다.

Vector를 matrix로 미분하면, 그 결과는 3차원 Vector이다.