CS7643 | Matrix Calculus Office Hour | 1/24/21

(2)

Scaler valued function of single var.

$f : \mathbb{R} \rightarrow \mathbb{R} \qquad x \in \mathbb{R}$

$$y = f(x) = x^2 \ln x^2 \Rightarrow \frac{dy}{dx} = 2x \ln x^2 + x^2 \frac{1}{x^2} 2x$$

$$= 2x \ln x^2 + 2x = 2x(\ln x^2 + 1)$$

---

Scaler valued func. of multi-vars

$f : \mathbb{R}^3 \rightarrow \mathbb{R} \qquad \vec{x} \in \mathbb{R}^3 \Rightarrow \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad \frac{\partial y}{\partial \vec{x}} \in \mathbb{R}^{1 \times 3}$

$\uparrow$
vector

$$y = f(\vec{x}) = x_1^2 + x_1 \ln x_2 + e^{x_2 x_3} + 2x_3$$

$$\frac{\partial y}{\partial x_1} = 2x_1 + \ln x_2$$

$$\frac{\partial y}{\partial x_2} = \frac{x_1}{x_2} + x_3 e^{x_2 x_3}$$

$$\frac{\partial y}{\partial x_3} = x e^{x_2 x_3} + 2$$

Numerator Layout : $\frac{\partial y}{\partial \vec{x}} = \left[ \frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2}, \frac{\partial y}{\partial x_3} \right] \in \mathbb{R}^{1 \times 3}$

what if $\vec{x} = [x_1\ x_2\ x_3] \rightarrow \frac{\partial y}{\partial \vec{x}} \in \mathbb{R}^{3 \times 1}$

---

(1)

Vector valued func. of single var

$f: \mathbb{R} \to \mathbb{R}^3 \qquad x \in \mathbb{R}$

$$\vec{y} = f(x) = \underbrace{\begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix}}_{\text{exist for all vals of } x} = \begin{bmatrix} 3x^2 + x \\ e^x \\ \ln x \end{bmatrix} \to \frac{\partial \vec{y}}{\partial x} = \begin{bmatrix} \frac{\partial f_1}{\partial x} \\ \frac{\partial f_2}{\partial x} \\ \frac{\partial f_3}{\partial x} \end{bmatrix} = \begin{bmatrix} 6x + 1 \\ e^x \\ \frac{1}{x} \end{bmatrix}$$

vector?

$$\frac{\partial \vec{y}}{\partial x} \in \mathbb{R}^3$$

---

Vector valued func. of multi-var

$$\vec{y} = \begin{bmatrix} f_1(\vec{x}) \\ f_2(\vec{x}) \end{bmatrix}$$

$f: \mathbb{R}^3 \to \mathbb{R}^2 \qquad \vec{x} \in \mathbb{R}^3 \Rightarrow \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad \vec{y} = f(\vec{x}) = \begin{bmatrix} x_1 + x_2 x_3 + x_1 x_3 \\ x_1^2 + 2x_3 \end{bmatrix}$

$$\frac{\partial \vec{y}}{\partial \vec{x}} \in \underbrace{\mathbb{R}^{2 \times 3}}_{\text{Jacobian}} \qquad \frac{\partial \vec{y}}{\partial \vec{x}} = \underbrace{\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \frac{\partial f_1}{\partial x_3} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \frac{\partial f_2}{\partial x_3} \end{bmatrix}}_{2 \times 3 \text{ matrix Jacobian}}$$

$\vec{f}$

$$\frac{\partial \vec{y}}{\partial \vec{x}} = \begin{bmatrix} 1 + x_3 & x_3 & x_2 + x_1 \\ 2x_1 & 0 & 2 \end{bmatrix}$$

Scaler valued func. of multi (matrix) vars

$f : \mathbb{R}^{3\times 4} \to \mathbb{R}$

$$X = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \\ x_5 & x_6 & x_7 & x_8 \\ x_9 & x_{10} & x_{11} & x_{12} \end{bmatrix}$$

$$y = f(X) = \sum_{i=1}^{3} \sum_{J=1}^{4} x_{iJ} \quad \text{(vector sum)}$$

$$\frac{\partial y}{\partial X} \in \mathbb{R}^{4\times 3} \Rightarrow \underbrace{\frac{\partial y}{\partial X}}_{\mathbb{R}^{4\times 3}} = \begin{bmatrix} \frac{\partial y}{\partial x_1} & \frac{\partial y}{\partial x_5} & \frac{\partial y}{\partial x_9} \\ \frac{\partial y}{\partial x_1} & \frac{\partial y}{\partial x_6} & \frac{\partial y}{\partial x_{10}} \\ \frac{\partial y}{\partial x_3} & \frac{\partial y}{\partial x_7} & \frac{\partial y}{\partial x_{11}} \\ \frac{\partial y}{\partial x_4} & \frac{\partial y}{\partial x_8} & \frac{\partial y}{\partial x_{12}} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

transpose in lecture $\Rightarrow$

Matrix valued func. w/ single var.

$f : \mathbb{R} \to \mathbb{R}^{2\times 2}$

$$y = f(x) = \begin{bmatrix} f_1(x) & f_2(x) \\ f_3(x) & f_4(x) \end{bmatrix} = \begin{bmatrix} \cos(x) & \sin(x) \\ e^x & \tanh(x) \end{bmatrix}$$

$$\frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_2}{\partial x} \\ \frac{\partial f_3}{\partial x} & \frac{\partial f_4}{\partial x} \end{bmatrix} = \begin{bmatrix} -\sin(x) & \cos(x) \\ e^x & 1 - \tanh^2(x) \end{bmatrix}$$

$\mathbb{R}^{2\times 3}$

$$\frac{\partial y}{\partial x} \in \mathbb{R}^{2\times 3}$$

Vector valued func. of multi (matrix) vars.

$$f: \mathbb{R}^{2\times 3} \to \mathbb{R}^4 \Rightarrow \text{3rd order tensor } 4\times 3\times 2$$

$$X = \begin{bmatrix} x_1 & x_2 & x_3 \\ x_4 & x_5 & x_6 \end{bmatrix}$$

$$\vec{y} = f(X) = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \end{bmatrix} = \begin{bmatrix} \sum_i \sum_j x_{ij} \\ x_1 x_5 x_6 \\ x_2 x_4 \\ x_3 x_4 x_5 \end{bmatrix}$$

$$J = \frac{\partial \vec{y}}{\partial X} = \underbrace{\begin{bmatrix} \frac{\partial \vec{y}}{\partial x_1} & \frac{\partial \vec{y}}{\partial x_4} \\ \frac{\partial \vec{y}}{\partial x_2} & \frac{\partial \vec{y}}{\partial x_5} \\ \frac{\partial \vec{y}}{\partial x_3} & \frac{\partial \vec{y}}{\partial x_6} \end{bmatrix}}_{\text{tensor}} =$$

$$\frac{\partial \vec{y}}{\partial x_1} = \begin{bmatrix} 1 \\ x_5 x_6 \\ 0 \\ 0 \end{bmatrix} \quad \frac{\partial \vec{y}}{\partial x_2} = \begin{bmatrix} 1 \\ 0 \\ x_4 \\ 0 \end{bmatrix}$$

$$\frac{\partial \vec{y}}{\partial x_3} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ x_4 x_5 \end{bmatrix} \quad \frac{\partial \vec{y}}{\partial x_4} = \begin{bmatrix} 1 \\ 0 \\ x_2 \\ x_3 x_5 \end{bmatrix}$$

$$\frac{\partial \vec{y}}{\partial x_5} = \begin{bmatrix} 1 \\ x_1 x_6 \\ 0 \\ x_3 x_4 \end{bmatrix} \quad \frac{\partial \vec{y}}{\partial x_6} = \begin{bmatrix} 1 \\ x_1 x_5 \\ 0 \\ 0 \end{bmatrix}$$

tensor $4\times 3\times 2$
$D \times R \times C$
depth row col

$$= \underbrace{\begin{bmatrix} 3\left\{\overbrace{\begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_4} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_1}{\partial x_5} \\ \frac{\partial y_1}{\partial x_3} & \frac{\partial y_1}{\partial x_6} \end{bmatrix}}^{2}\right. & \begin{bmatrix} \frac{\partial y_2}{\partial X} \end{bmatrix} & \begin{bmatrix} \frac{\partial y_3}{\partial X} \end{bmatrix} & \begin{bmatrix} \frac{\partial y_4}{\partial X} \end{bmatrix} \end{bmatrix}}_{4}$$

$$
= \left[ \underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}} \begin{bmatrix} x_5 x_6 & 0 \\ 0 & x_1 x_6 \\ 0 & x_1 x_5 \end{bmatrix} \begin{bmatrix} 0 & x_2 \\ x_4 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & x_3 x_5 \\ 0 & x_3 x_4 \\ x_4 x_5 & 0 \end{bmatrix} \right]
$$

$$
\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_4} \\ \frac{\partial f_1}{\partial x_2} & \frac{\partial f_1}{\partial x_5} \\ \frac{\partial f_1}{\partial x_3} & \frac{\partial f_1}{\partial x_6} \end{bmatrix}
$$

flatten Jacobian: $f : \mathbb{R}^{2\times 3} \to \mathbb{R}^4$ $\quad X \in \mathbb{R}^{2\times 3}$

$\Downarrow$ $\quad\quad \Downarrow$ vectorize

$f : \mathbb{R}^6 \to \mathbb{R}^4$ $\quad vec(X) = [x_1\, x_2\, x_3\, x_4\, x_5\, x_6]^T$

- flatten out into vec. to avoid tensors: ① vectorize ② sparse J $\to$ vec
- In batch, it's all matrix math

derivative of I (identity matrix)

$IX = X \quad X \in \mathbb{R}^{n \times m}$

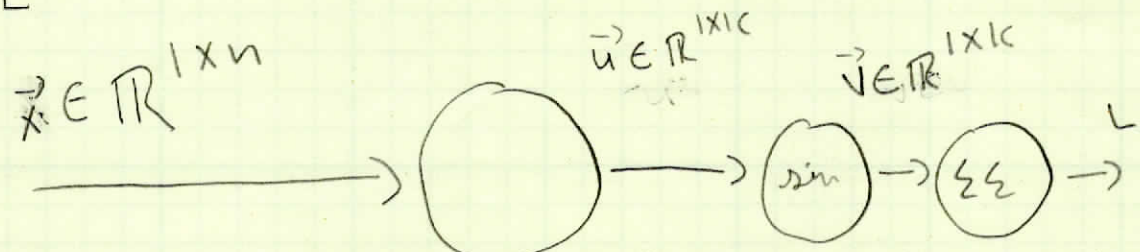
$f : \mathbb{R}^3 \to \mathbb{R}^3 \qquad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

$$\vec{y} = f(\vec{x}) = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$\frac{\partial \vec{y}}{\partial \vec{x}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \frac{\partial f_1}{\partial x_3} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \frac{\partial f_2}{\partial x_3} \\ \frac{\partial f_3}{\partial x_1} & \frac{\partial f_3}{\partial x_2} & \frac{\partial f_3}{\partial x_3} \end{bmatrix}$$

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \Rightarrow [1\ 1\ 1]^T$$

$\to$ sparse J (use diag. non-zero as vec.)




$W \in \mathbb{R}^{n \times k}$

$$\frac{\partial L}{\partial v} = \vec{1} \in \mathbb{R}^{1 \times k}$$

$$\frac{\partial v}{\partial u} = \cos(\vec{u}) \in \mathbb{R}^{k \times k}$$

$$\frac{\partial L}{\partial W} = \frac{\partial L}{\partial v} \cdot \frac{\partial v}{\partial u} \cdot \frac{\partial u}{\partial W}$$

$1 \times k \quad k \times k \quad 1 \times n$

$1 \times k \quad 1 \times n$

$$\frac{\partial u}{\partial W} = \frac{\partial}{\partial W}(\vec{x}W) = \vec{x} \in \mathbb{R}^{1 \times n}$$

$$\left[\left[(1\times k)\underbrace{(k\times k)}\right]^T \times (1\times n)\right]^T \in \mathbb{R}^{n\times k}$$

$$\Downarrow$$

sparsity

$$(1\times k)^T \Rightarrow (k\times 1)(1\times n) = (k\times n)^T \Rightarrow n\times k$$

hadamard product (element-wise product)

$$(1\times k)\circ(1\times k)$$

$$(1\times k) = \mathrm{diag}(k\times k)$$

---

$$\vec{x}\in\mathbb{R}^{1\times n} \longrightarrow \bigcirc \xrightarrow{\vec{u}\in\mathbb{R}^{1\times k}} (\sin) \xrightarrow{\vec{v}\in\mathbb{R}^{1\times k}} \left(\sum_i \sum_J\right) \longrightarrow L\in\mathbb{R}$$

$$W\in\mathbb{R}^{n\times k} \uparrow$$

$\vec{u}$?

$$\frac{\partial \vec{v}}{\partial \vec{u}} = \cos(\vec{u})$$

$\vec{u} ?= W^T x \rightarrow$ stored somewhere

input $\longrightarrow$ (soft) $\xrightarrow{+}$ (CE) $\xrightarrow{\div}$ L

combine

know when func: is differentiable (poly)

subgradient?? is piecewise diff (abs. of ReLU)

is non-diff (unit step)

??

ln
$e^x$
sin
cos
tan
tanh
$\sigma(x)$

$$\sigma'(x) = \frac{d}{dx}\left(\frac{1}{1+e^{-x}}\right)$$

$$= \frac{d}{dx}(1+e^{-x})^{-1}$$

$$= -(1+e^{-x})^{-2} \cdot e^{-x} \cdot -1$$

$$= e^{-x}(1+e^{-x})^{-2}$$

$$= \sigma(x) \cdot \frac{e^{-x}}{(1+e^{-x})}$$

$$\frac{e^{-x}}{1+e^{-x}} = \frac{1+e^{-x}-1}{1+e^{-x}} = \frac{1+e^{-x}}{1+e^{-x}} - \frac{1}{1+e^{-x}}$$

$$\Downarrow$$

$$\sigma(x)\left[\overset{1}{\cancel{\frac{1+e^{-x}}{1+e^{-x}}}} - \overset{\sigma}{\cancel{\frac{1}{1+e^{-x}}}}\right]$$

$$\sigma'(x) = \sigma(x)\left(1 - \sigma(x)\right)$$