Matrix Calculus Examples

Farrukh Rahman

Agenda

- Prelim
- Examples
- Num vs Denom Layout
- NN

Why?

$$f(x, \theta) = y$$

L

- NN's approximate functions
- Gradient is response to infinitesimal changes.
    - Gradient based Learning
- Backpropagation workhorse of DL

functions

- f: X -> Y
  - Function f maps X to Y
  - Input X can be Scalar, Vector, Matrix
  - Output Y can be Scalar, Vector, or Matrix
  - Vectors are column vectors.

Scalar Valued Function of Single Variable

$f: \mathbb{R} \to \mathbb{R}$ $\quad x \in \mathbb{R}$

$y = f(x) = x^2 \ln x^2$

$$\frac{dy}{dx} = x^2 \frac{2x}{x^2} + 2x \ln x^2$$

Scalar Valued Function of Multiple variables

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

$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

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

$y = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = \begin{bmatrix} \text{—} \cdots \end{bmatrix}$

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

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

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

[denominator] $\leftrightarrow$ [numerator]

$\frac{\partial y}{\partial x} \in \mathbb{R}^{1\times 3}$

$\begin{bmatrix} \frac{\partial y}{\partial x_1} & \frac{\partial y}{\partial x_2} & \frac{\partial y}{\partial x_3} \end{bmatrix}$

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

Vector Valued Function of Single Variable

$f : \mathbb{R} \to \mathbb{R}^3$ $\quad y = f(x) = \begin{bmatrix} 3x^2 + x \\ e^x \\ \ln x \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix}$ $\quad x \in \mathbb{R}$

$$\frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial f_1}{\partial x} \\ \frac{\partial f_2}{\partial x} \\ \frac{\partial f_3}{\partial x} \end{bmatrix} = \begin{bmatrix} 6x + 1 \\ e^x \\ 1/x \end{bmatrix}$$

Vector Valued Function of Multiple Variables (numerator layout)

$f: \mathbb{R}^3 \to \mathbb{R}^2$ $\qquad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

$$y = f(\vec{x}) = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = \begin{bmatrix} x_1 + x_1 x_2 + x_1 x_3 \\ x_1^2 + 2x_3 \end{bmatrix}$$

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

Jacobian

$$= \begin{bmatrix} 1 + x_2 + x_3 & x_1 & x_1 \\ 2x_1 & 0 & 2 \end{bmatrix} \in \mathbb{R}^{2 \times 3}$$

Vector Valued Function of Multiple Variables (denominator layout)

$f: \mathbb{R}^3 \to \mathbb{R}^2$ $\qquad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

$$y = f(\vec{x}) = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = \begin{bmatrix} x_1 + x_1 x_2 + x_1 x_3 \\ x_1^2 + 2x_3 \end{bmatrix}$$

$$\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} \\ \frac{\partial f_1}{\partial x_2} & \frac{\partial f_2}{\partial x_2} \\ \frac{\partial f_1}{\partial x_3} & \frac{\partial f_2}{\partial x_3} \end{bmatrix}$$

Scalar valued function of many (matrix) variables

$$f : \mathbb{R}^{3\times4} \to \mathbb{R}$$

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

$$y = f(X) = \sum_i \sum_j x_{ij}$$

$$\frac{\partial y}{\partial X} = \begin{bmatrix} \frac{\partial y}{\partial x_1} & \frac{\partial y}{\partial x_5} & \frac{\partial y}{\partial x_9} \\ \frac{\partial y}{\partial x_2} & \frac{\partial y}{\partial x_6} & \frac{\partial y}{\partial x_{10}} \\ \frac{\partial y}{\partial x_3} & \frac{\partial y}{\partial x_7} & \frac{\partial y}{\partial x_{11}} \\ \frac{\partial y}{\partial x_4} & \frac{\partial y}{\partial x_8} & \frac{\partial y}{\partial x_{12}} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \in \mathbb{R}^{4\times3}$$

Matrix valued function of single variable

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

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

Vector valued function of multiple (matrix) variables pt 1.

$$f : \mathbb{R}^{2\times 3} \to \mathbb{R}^{4} \quad (\mathbb{R}^{1})$$

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

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

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

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

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

$$\begin{bmatrix} ① \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \\ ② \begin{bmatrix} x_5 x_6 & 0 \\ 0 & x_1 x_6 \\ 0 & x_1 x_5 \end{bmatrix} \\ ③ \begin{bmatrix} 0 & x_2 \\ x_4 & 0 \\ 0 & 0 \end{bmatrix} \\ ④ \begin{bmatrix} 0 & x_3 x_5 \\ 0 & x_3 x_4 \\ x_4 x_5 & 0 \end{bmatrix} \end{bmatrix} \in \mathbb{R}^{4\times 3\times 2}$$

$$\begin{bmatrix} ① \\ ② \\ ③ \\ ④ \end{bmatrix}$$

Vector valued function of multiple (matrix) variables pt 2.

(vectorized) $f : \mathbb{R}^{2\times 3} \to \mathbb{R}^4$

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

$$\mathrm{Vec}(x) = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \end{bmatrix}^T \in \mathbb{R}^6$$

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

$$\frac{\partial y}{\partial \mathrm{vec}(x)} \in \mathbb{R}^{4\times 6}$$

$$\begin{bmatrix} \quad \end{bmatrix}$$

Matrix valued function of multiple variables

$f : \mathbb{R}^3 \to \mathbb{R}^{2\times 4}$ $\quad \vec{x} = [x_1 \; x_2 \; x_3]^T$ $\quad y = f(x) = \begin{bmatrix} \sum_i x_i & x_1 x_3 & x_2 x_3 & x_2 \\ x_1 x_2 & x_1^2 & x_2^2 & x_3 \end{bmatrix}$

$\frac{\partial y}{\partial \vec{x}} \in \mathbb{R}^{2\times 4\times 3}$

Elementwise functions (e.g. Identity) $f: \mathbb{R}^{n\times k} \to \mathbb{R}^{n\times k}$

$f: \mathbb{R}^{3\times 1} \to \mathbb{R}^{3\times 1}$ $\quad \vec{x} = [x_1 \ x_2 \ x_3]^T$

$I(x) = x \in \mathbb{R}$

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

$f(x) = x$

$\frac{df(x)}{dx} = 1$

$f_1(x_1) = x_1$

$f_2(x_2) = x_2$

$f_3(x_3) = x_3$

$$\frac{\partial y}{\partial \vec{x}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \frac{\partial f_1}{\partial x_3} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \frac{\partial f_2}{\partial x_3} \\ \frac{\partial f_3}{\partial x_1} & \frac{\partial f_3}{\partial x_2} & \frac{\partial f_3}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 1 \ (\cos x_1) & 0 & 0 \\ 0 & \cos x_2 & 0 \\ 0 & 0 & 1 \ (\cos x_3) \end{bmatrix}$$

$f_1 = \sin x_1$

$f_2 = \sin x_2$

$f_3 = \sin x_3$

$f(X)$

$$\begin{bmatrix} f(x_{11}) & f(x_{12}) \\ f(x_{21}) & f(x_{22}) \end{bmatrix}$$

Layout notation.

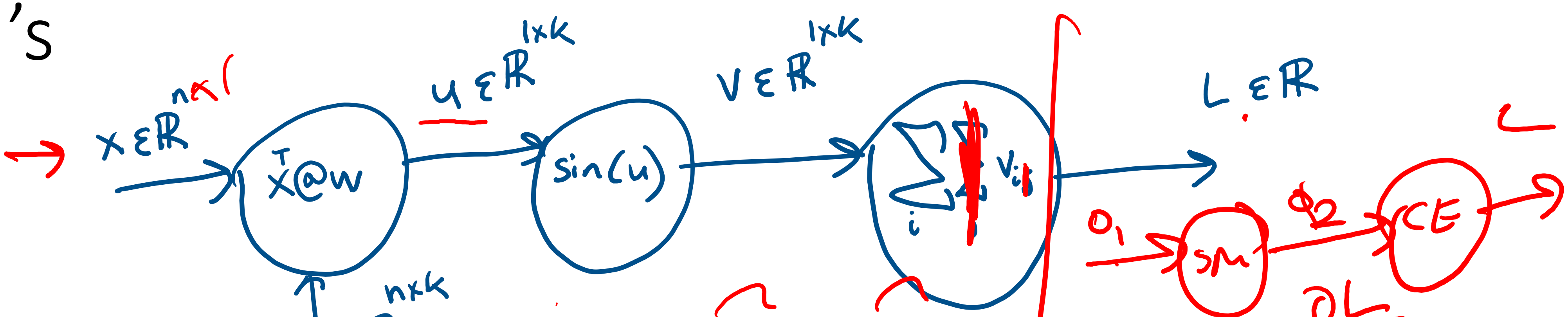
- Use whichever layout is convenient to make shapes match.
- Why? Update rule

Result of differentiating various kinds of aggregates with other kinds of aggregates

		Scalar y		Column vector $\mathbf{y}$ (size m×1)		Matrix $\mathbf{Y}$ (size m×n)	
		Notation	Type	Notation	Type	Notation	Type
Scalar x	Numerator	$\frac{\partial y}{\partial x}$	Scalar	$\frac{\partial \mathbf{y}}{\partial x}$	Size-m column vector	$\frac{\partial \mathbf{Y}}{\partial x}$	m×n matrix
	Denominator				Size-m row vector		
Column vector $\mathbf{x}$ (size n×1)	Numerator	$\frac{\partial y}{\partial \mathbf{x}}$	Size-n row vector	$\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$	m×n matrix	$\frac{\partial \mathbf{Y}}{\partial \mathbf{x}}$	
	Denominator		Size-n column vector		n×m matrix		
Matrix $\mathbf{X}$ (size p×q)	Numerator	$\frac{\partial y}{\partial \mathbf{X}}$	q×p matrix	$\frac{\partial \mathbf{y}}{\partial \mathbf{X}}$		$\frac{\partial \mathbf{Y}}{\partial \mathbf{X}}$	
	Denominator		p×q matrix				

The results of operations will be transposed when switching between numerator-layout and denominator-layout notation.

Back to NN's



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

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

$$= \left[(1\times k)\odot(1\times k)\right]^T (n\times 1)^T$$

$(1\times k)$ $\quad n\times 1$ $\quad n\times k$

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

$$\frac{\partial v}{\partial u} \in \mathbb{R}^{k\times k} \quad \text{full jacobian} \qquad \in \mathbb{R}^{1\times k} \quad \text{diag}(J)$$

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