Agenda

- Prelim
- Examples
- ~~Num vs Denom Layout~~

Why?

$f(\underline{x})$

- Derivative is response to infinitesimal changes in input.
- Backpropagation workhorse of DL

Prelim

$f: X \to Y$

$x \to Y$

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

Scalar Valued Function of Single Variable

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

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

$$\frac{\partial y}{\partial x} = x^2 \frac{2x}{x^2} + 2x \ln x^2 = 2x(1 + \ln x^2)$$

Scalar Valued Function of Multiple variables

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

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

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

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

$$\begin{bmatrix} \cancel{\text{numerator}} \end{bmatrix} \in \mathbb{R}^{1\times 3}$$

$$\begin{bmatrix} \cancel{\begin{matrix} d \\ e \\ n \\ o \\ m \end{matrix}} \end{bmatrix} \in \mathbb{R}^{3\times 1}$$

Vector Valued Function of Single Variable

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

$$\frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial (3x^2 + x)}{\partial x} \\ \frac{\partial (e^x)}{\partial x} \\ \frac{\partial (\ln x)}{\partial x} \end{bmatrix} = \begin{bmatrix} 6x + 1 \\ e^x \\ \frac{1}{x} \end{bmatrix} \in \mathbb{R}^{3 \times 1}$$

Vector Valued Function of Multiple Variables (numerator layout)

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

$$y = f(\vec{x}) = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = \begin{bmatrix} x_1 + x_1x_2 + x_1x_3 \\ x_1^2 + 2x_3 \end{bmatrix}$$

$$\frac{\partial f(\vec{x})}{\partial \vec{x}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \frac{\partial f_1}{\partial x_3} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \frac{\partial f_2}{\partial x_3} \end{bmatrix} \qquad \text{Jacobian}$$

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

Vector Valued Function of Multiple Variables (denominator layout)

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

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

$$\frac{\partial f(\vec{x})}{\partial \vec{x}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} \\ \frac{\partial f_1}{\partial x_2} & \frac{\partial f_2}{\partial x_2} \\ \frac{\partial f_1}{\partial x_3} & \frac{\partial f_2}{\partial x_3} \end{bmatrix} \in \mathbb{R}^{3\times 2}$$

Scalar valued function of many (matrix) variables

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

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ x_{31} & x_{32} & x_{33} & x_{34} \end{bmatrix}$$

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

$$\frac{\partial y}{\partial X} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{21}} & \frac{\partial y}{\partial x_{31}} \\ \frac{\partial y}{\partial x_{12}} & \frac{\partial y}{\partial x_{22}} & \frac{\partial y}{\partial x_{32}} \\ \frac{\partial y}{\partial x_{13}} & \frac{\partial y}{\partial x_{23}} & \frac{\partial y}{\partial x_{33}} \\ \frac{\partial y}{\partial x_{14}} & \frac{\partial y}{\partial x_{24}} & \frac{\partial y}{\partial x_{34}} \end{bmatrix} = \begin{bmatrix} | & | & | \\ | & | & | \\ | & | & | \\ | & | & | \end{bmatrix} \in \mathbb{R}^{4\times 3}$$

Matrix valued function of single variable

$x = 0$

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

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

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

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

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

$\in \mathbb{R}^{2\times3\times4}$

$$f : \mathbb{R}^{2\times3} \to \mathbb{R}^4 \qquad x = \begin{bmatrix} x_1 & x_2 & x_3 \\ x_4 & x_5 & x_6 \end{bmatrix}$$

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

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

$$\frac{\partial y}{\partial x_1} = \begin{bmatrix} 1 \\ x_5 x_6 \\ 0 \\ 0 \end{bmatrix} \quad \frac{\partial y}{\partial x_2} = \begin{bmatrix} 1 \\ 0 \\ x_4 \\ 0 \end{bmatrix} \quad \frac{\partial y}{\partial x_3} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ x_4 x_5 \end{bmatrix}$$

$$\frac{\partial y}{\partial x_4} = \begin{bmatrix} 1 \\ 0 \\ x_2 \\ x_3 x_5 \end{bmatrix} \quad \frac{\partial y}{\partial x_5} = \begin{bmatrix} 1 \\ x_1 x_6 \\ 0 \\ x_3 x_4 \end{bmatrix} \quad \frac{\partial y}{\partial x_6} = \begin{bmatrix} 1 \\ x_1 x_5 \\ 0 \\ 0 \end{bmatrix}$$

$$\left[\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_5 x_6 & 0 \\ 0 & x_1 x_6 \\ 0 & x_1 x_5 \end{bmatrix} \begin{bmatrix} 0 & x_2 \\ x_4 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & x_3 x_5 \\ 0 & x_3 x_4 \\ x_4 x_5 & 0 \end{bmatrix} \right] \in \mathbb{R}^{4\times3\times2}$$

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

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

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

$$\operatorname{vec}(X) = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \end{bmatrix}^T \in \mathbb{R}^6$$

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

$$\boxed{\frac{\partial y}{\partial \operatorname{vec}(X)} \in \mathbb{R}^{4\times 6}}$$

Matrix valued function of multiple variables

$f: \underline{\mathbb{R}^3} \to \underline{\mathbb{R}^{2\times 4}}$

$\vec{x} = [x_1 \; x_2 \; x_3]^T$

$y = f(x) = \begin{bmatrix} \sum_i x_i & x_1 x_3 & x_2 x_3 & x_2 \\ x_1 x_2 & x_1^2 & x_2^2 & x_3 \end{bmatrix}$

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

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

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

3x3

$$y = f(\vec{x}) = \begin{bmatrix} I_1(x_1) \\ I_2(x_2) \\ I_3(x_3) \end{bmatrix} \qquad I_1(x) = x \in \mathbb{R}$$

$\to [1 \ 1 \ 1]$

$$\frac{\partial y}{\partial \vec{x}} = \begin{bmatrix} \frac{\partial I_1}{\partial x_1} & \frac{\partial I_1}{\partial x_2} & \frac{\partial I_1}{\partial x_3} \\ \frac{\partial I_2}{\partial x_1} & \frac{\partial I_2}{\partial x_2} & \frac{\partial I_2}{\partial x_3} \\ \frac{\partial I_3}{\partial x_1} & \frac{\partial I_3}{\partial x_2} & \frac{\partial I_3}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Layout notation.

- Often sources will 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 $m \times 1$)**		**Matrix $\mathbf{Y}$ (size $m \times n$)**	
		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	$\frac{\partial \mathbf{Y}}{\partial x}$	$m \times n$ matrix
	Denominator				Size-m row vector		
Column vector $\mathbf{x}$ (size $n \times 1$)	**Numerator**	$\frac{\partial y}{\partial \mathbf{x}}$	Size-n row vector	$\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$	$m \times n$ matrix	$\frac{\partial \mathbf{Y}}{\partial \mathbf{x}}$	
	Denominator		Size-n column vector		$n \times m$ matrix		
Matrix $\mathbf{X}$ (size $p \times q$)	**Numerator**	$\frac{\partial y}{\partial \mathbf{X}}$	$q \times p$ matrix	$\frac{\partial \mathbf{y}}{\partial \mathbf{X}}$		$\frac{\partial \mathbf{Y}}{\partial \mathbf{X}}$	
	Denominator		$p \times q$ matrix				

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

Source: wiki

Conventions:

- Size of derivatives for scalars, vectors, and matrices: Assume we have scalar $s \in \mathbb{R}^1$, vector $v \in \mathbb{R}^m$, i.e. $v = [v_1, v_2, \dots, v_m]^T$ and matrix $M \in \mathbb{R}^{k \times \ell}$

numerator

$$\begin{bmatrix} \frac{\partial v_1}{\partial s} \\ \frac{\partial v_2}{\partial s} \\ \vdots \\ \frac{\partial v_m}{\partial s} \end{bmatrix}$$

- What is the size of $\frac{\partial v}{\partial s}$? $\mathbb{R}^{m \times 1}$ (column vector of size m)
- What is the size of $\frac{\partial s}{\partial v}$? $\mathbb{R}^{1 \times m}$ (row vector of size m)

$$\left[\frac{\partial s}{\partial v_1} \; \frac{\partial s}{\partial v_1} \; \cdots \; \frac{\partial s}{\partial v_m}\right]$$

Conventions:

- What is the size of $\frac{\partial v^1}{\partial v^2}$? A matrix:

numerator

$$
\begin{array}{cc}
 & \text{Col } j \\
\text{Row } i & \begin{bmatrix}
\frac{\partial v^1_1}{\partial v^2_1} & \cdots & \cdots & \cdots & \cdots & \\
\cdots & & \cdots & \cdots & \cdots & \cdots \\
\cdots & \cdots & \frac{\partial v^1_i}{\partial v^2_j} & \cdots & \cdots & \\
\cdots & \cdots & \cdots & \cdots & \cdots & \\
\cdots & \cdots & \cdots & \cdots & \cdots &
\end{bmatrix}
\end{array}
$$

- This matrix of partial derivatives is called a **Jacobian**

$\frac{\partial L}{\partial w}$ $\frac{s}{m}$ $\rightarrow$ $q \times p$

Conventions:

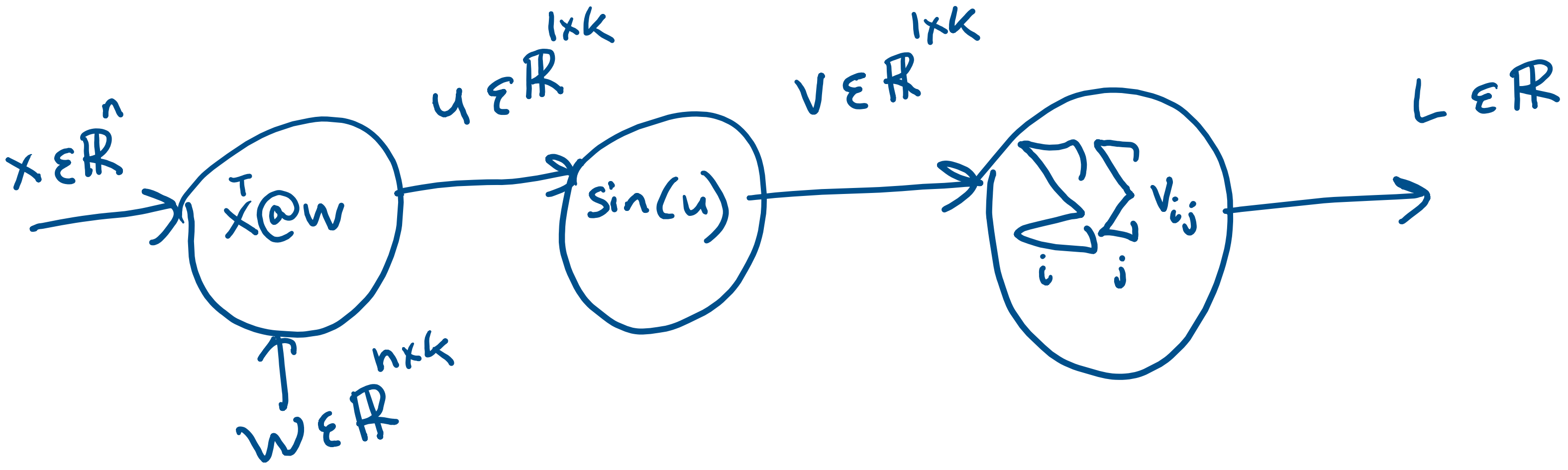
- What is the size of $\frac{\partial s}{\partial M}$? A matrix:

denominator

$$\begin{bmatrix} \frac{\partial s}{\partial m_{[1,1]}} & \cdots & \cdots & \cdots & \cdots & \\ \cdots & & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \frac{\partial s}{\partial m_{[i,j]}} & \cdots & \cdots & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \end{bmatrix}$$

Extra slides

Back to NN's



$x \in \mathbb{R}^n$
$x^T @ W$
$W \in \mathbb{R}^{n \times k}$
$u \in \mathbb{R}^{1 \times k}$
$\sin(u)$
$v \in \mathbb{R}^{1 \times k}$
$\sum_i \sum_j v_{ij}$
$L \in \mathbb{R}$