IT5005 Artificial Intelligence

Matrix Calculus

Sirigina Rajendra Prasad

Matrix Calculus (Denominator Layout)

- Let $\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_r \end{bmatrix}_{r\times 1}$ $\quad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}_{n\times 1}$ $\quad \boldsymbol{X} = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix}_{m\times n}$

$$\frac{\partial y}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix}_{n\times 1}$$

$$\frac{\partial y}{\partial \boldsymbol{X}} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \cdots & \frac{\partial y}{\partial x_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial y}{\partial x_{m1}} & \cdots & \frac{\partial y}{\partial x_{mn}} \end{bmatrix}_{m\times n}$$

$$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_r}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial x_n} & \cdots & \frac{\partial y_r}{\partial x_n} \end{bmatrix}_{n\times r}$$

$$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{X}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_{11}} & \cdots & \frac{\partial y_1}{\partial x_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial x_{m1}} & \cdots & \frac{\partial y_1}{\partial x_{mn}} \end{bmatrix} \ddots \begin{bmatrix} \frac{\partial y_r}{\partial x_{11}} & \cdots & \frac{\partial y_r}{\partial x_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_r}{\partial x_{m1}} & \cdots & \frac{\partial y_r}{\partial x_{mn}} \end{bmatrix}_{m\times n\times r}$$

Matrix Calculus (Denominator Layout)

- Vector-by-Vector Identities:

Condition	Expression	Numerator layout, i.e. by $\mathbf{y}$ and $\mathbf{x}^{\mathrm{T}}$	Denominator layout, i.e. by $\mathbf{y}^{\mathrm{T}}$ and $\mathbf{x}$
$\mathbf{a}$ is not a function of $\mathbf{x}$	$\frac{\partial \mathbf{a}}{\partial \mathbf{x}} =$	$\mathbf{0}$	
	$\frac{\partial \mathbf{x}}{\partial \mathbf{x}} =$	$\mathbf{I}$	
$\mathbf{A}$ is not a function of $\mathbf{x}$	$\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} =$	$\mathbf{A}$	$\mathbf{A}^\top$
$\mathbf{A}$ is not a function of $\mathbf{x}$	$\frac{\partial \mathbf{x}^\top \mathbf{A}}{\partial \mathbf{x}} =$	$\mathbf{A}^\top$	$\mathbf{A}$
a is not a function of $\mathbf{x}$, $\mathbf{u} = \mathbf{u}(\mathbf{x})$	$\frac{\partial a \mathbf{u}}{\partial \mathbf{x}} =$	$a \frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	
$v = v(\mathbf{x})$, $\mathbf{a}$ is not a function of $\mathbf{x}$	$\frac{\partial v \mathbf{a}}{\partial \mathbf{x}} =$	$\mathbf{a} \frac{\partial v}{\partial \mathbf{x}}$	$\frac{\partial v}{\partial \mathbf{x}} \mathbf{a}^\top$
$v = v(\mathbf{x})$, $\mathbf{u} = \mathbf{u}(\mathbf{x})$	$\frac{\partial v \mathbf{u}}{\partial \mathbf{x}} =$	$v \frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \mathbf{u} \frac{\partial v}{\partial \mathbf{x}}$	$v \frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \frac{\partial v}{\partial \mathbf{x}} \mathbf{u}^\top$
$\mathbf{A}$ is not a function of $\mathbf{x}$, $\mathbf{u} = \mathbf{u}(\mathbf{x})$	$\frac{\partial \mathbf{A}\mathbf{u}}{\partial \mathbf{x}} =$	$\mathbf{A} \frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} \mathbf{A}^\top$
$\mathbf{u} = \mathbf{u}(\mathbf{x})$, $\mathbf{v} = \mathbf{v}(\mathbf{x})$	$\frac{\partial (\mathbf{u} + \mathbf{v})}{\partial \mathbf{x}} =$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \frac{\partial \mathbf{v}}{\partial \mathbf{x}}$	
$\mathbf{u} = \mathbf{u}(\mathbf{x})$	$\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{x}} =$	$\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}} \frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} \frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}$
$\mathbf{u} = \mathbf{u}(\mathbf{x})$	$\frac{\partial \mathbf{f}(\mathbf{g}(\mathbf{u}))}{\partial \mathbf{x}} =$	$\frac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{g}} \frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}} \frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} \frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}} \frac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{g}}$

https://en.wikipedia.org/wiki/Matrix_calculus

Chain Rule (Denominator Layout)

- $\boldsymbol{g}$ is a vector-valued function with output vector of shape $m \times 1$
- $\boldsymbol{f}$ is a column vector of shape $d \times 1$
- $\boldsymbol{x}$ is a column vector of shape $n \times 1$

$$\underbrace{\frac{\partial \boldsymbol{g}(\boldsymbol{f})}{\partial \boldsymbol{x}}}_{n \times m} = \underbrace{\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}}_{n \times d} \underbrace{\frac{\partial \boldsymbol{g}(\boldsymbol{f})}{\partial \boldsymbol{f}}}_{d \times m}$$

Matrix Calculus: Matrix Identities

- Let $\boldsymbol{z} = W\boldsymbol{x}$, where $\boldsymbol{x} \in \mathbb{R}^{m\times 1}, W \in \mathbb{R}^{n\times m}$. Find $\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}}$?

$$\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial z_1}{\partial x_1} & \cdots & \frac{\partial z_n}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial z_1}{\partial x_m} & \cdots & \frac{\partial z_n}{\partial x_m} \end{bmatrix}$$

$$z_i = \sum_{k=1}^{m} W_{ik} x_k \Longrightarrow \frac{\partial z_i}{\partial x_k} = W_{ik}$$

$$\Longrightarrow \frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}} = \begin{bmatrix} W_{11} & \cdots & W_{n1} \\ \vdots & \ddots & \vdots \\ W_{1m} & \cdots & W_{nm} \end{bmatrix} = W^T$$

Matrix Calculus: Matrix Identities

- Let $\boldsymbol{z} = \boldsymbol{x}^T W$, where $\boldsymbol{x} \in \mathbb{R}^{m\times 1}, W \in \mathbb{R}^{m\times n}$. Find $\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}}$?

$$\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial z_1}{\partial x_1} & \cdots & \frac{\partial z_n}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial z_1}{\partial x_m} & \cdots & \frac{\partial z_n}{\partial x_m} \end{bmatrix}$$

$$z_i = \sum_{k=1}^{m} x_k W_{ki} \Longrightarrow \frac{\partial z_i}{\partial x_j} = W_{ji}$$

$$\Longrightarrow \frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}} = \begin{bmatrix} W_{11} & \cdots & W_{1n} \\ \vdots & \ddots & \vdots \\ W_{m1} & \cdots & W_{mn} \end{bmatrix} = W$$

Matrix Calculus: Matrix Identities

- Let $\boldsymbol{x} \in \mathbb{R}^{m\times 1}$ and $\mathbf{z} = g(\boldsymbol{x}) = \begin{bmatrix} g(x_1) \\ \vdots \\ g(x_m) \end{bmatrix}, i.e.,$ function g performs element-wise application for the vector $\boldsymbol{x}$. Find $\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}}$?

$$\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial z_1}{\partial x_1} & \cdots & \frac{\partial z_m}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial z_1}{\partial x_m} & \cdots & \frac{\partial z_m}{\partial x_m} \end{bmatrix} = \begin{bmatrix} \frac{\partial g(x_1)}{\partial x_1} & \cdots & \frac{\partial g(x_m)}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial g(x_1)}{\partial x_m} & \cdots & \frac{\partial g(x_1)}{\partial x_m} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\partial g(x_1)}{\partial x_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\partial g(x_1)}{\partial x_m} \end{bmatrix} = \boldsymbol{diag}[z']$$

$$z' = g'(\boldsymbol{x}) = \begin{bmatrix} g'(x_1) \\ \vdots \\ g'(x_m) \end{bmatrix}$$

Matrix Calculus: Matrix Identities

- Let $\boldsymbol{f} = W^T\boldsymbol{a}$, where $W \in \mathbb{R}^{m\times n}$ and $\boldsymbol{a} \in \mathbb{R}^{m\times 1}$. Let $g\big(f(\boldsymbol{W})\big)$ be a scalar function. Assuming denominator layout show that

$$\frac{\partial g(\boldsymbol{f}(W))}{\partial W} = \frac{\partial \boldsymbol{f}(W)}{\partial W} \quad \frac{\partial g(\boldsymbol{f})}{\partial \boldsymbol{f}} = \boldsymbol{a}\left(\frac{\partial g(\boldsymbol{f})}{\partial \boldsymbol{f}}\right)^T$$

Chain Rule in Denominator Layout:

$$\underbrace{\frac{\partial g(\boldsymbol{f}(W))}{\partial W}}_{m\times n} = \underbrace{\frac{\partial \boldsymbol{f}(W)}{\partial W}}_{m\times n\times n} \underbrace{\frac{\partial g(\boldsymbol{f})}{\partial \boldsymbol{f}}}_{n\times 1}$$

$\boldsymbol{a} \in \mathbb{R}^{m\times 1}$

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

$\boldsymbol{f} \in \mathbb{R}^{n\times 1}$

$g(\boldsymbol{f})\colon\ \mathbb{R}^{n\times 1} \to \mathbb{R}$

- Proof:
  - Let's start with $\frac{\partial g(\boldsymbol{f}(W))}{\partial W_{ij}}$, where W_{ij} is a scalar from i-th row and j-th column of matrix W

$$\underbrace{\frac{\partial g(\boldsymbol{f})}{\partial W_{ij}}}_{1\times 1} = \underbrace{\frac{\partial \boldsymbol{f}(W)}{\partial W_{ij}}}_{1\times n} \underbrace{\frac{\partial g(\boldsymbol{f}))}{\partial \boldsymbol{f}}}_{n\times 1}$$

$\boldsymbol{a} \in \mathbb{R}^{m\times 1}$

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

$\boldsymbol{f} \in \mathbb{R}^{n\times 1}$

$g(\boldsymbol{f})$: $\mathbb{R}^{n\times 1} \rightarrow \mathbb{R}$

- Proof (contd):

$$\frac{\partial g(\boldsymbol{f}(W))}{\partial W_{ij}} = \frac{\partial \boldsymbol{f}}{\partial W_{ij}} \frac{\partial g(\boldsymbol{f}(W))}{\partial \boldsymbol{f}}$$

$$\frac{\partial \boldsymbol{f}(\boldsymbol{W})}{\partial W_{ij}} = \left[\frac{\partial f_1(W)}{\partial W_{ij}} \quad \cdots \quad \frac{\partial f_k(W)}{\partial W_{ij}} \quad \cdots \quad \frac{\partial f_n(W)}{\partial W_{ij}}\right] \in \mathbb{R}^{1\times n}$$

$$\boldsymbol{f} = W^T \boldsymbol{a}$$

k-th column of matrix W

$$\Longrightarrow \; f_k = \sum_{l=1}^{m} W_{lk} a_l$$

$$\frac{\partial f_k(W)}{\partial W_{ij}} = \frac{\partial}{\partial W_{ij}} \left(\sum_{l=1}^{m} W_{lk} a_l \right)$$

$$= \begin{cases} a_i & if \;\; l = i \; and \; k = j \\ 0 & otherwise \end{cases}$$

$$\Longrightarrow \frac{\partial \boldsymbol{f}(W)}{\partial W_{ij}} = [0 \quad \cdots \quad 0 \quad a_i \quad 0 \quad \ldots \quad 0]$$

$$\Downarrow$$

$j - th$ element of the vector $\frac{\partial \boldsymbol{f}(W)}{\partial W_{ij}}$

$\boldsymbol{a} \in \mathbb{R}^{m\times 1}$
$W \in \mathbb{R}^{m\times n}$
$\boldsymbol{f} \in \mathbb{R}^{n\times 1}$
$g(\boldsymbol{f})\colon \mathbb{R}^{n\times 1} \to \mathbb{R}$

f_k: k-th element of vector $\boldsymbol{f}$

- Proof (contd):

$$\frac{\partial g(\boldsymbol{f}(W))}{\partial W_{ij}} = \frac{\partial \boldsymbol{f}(W)}{\partial W_{ij}} \frac{\partial g(\boldsymbol{f})}{\partial \boldsymbol{f}}$$

Let $\boldsymbol{\delta} = \frac{\partial g(\boldsymbol{f})}{\partial \boldsymbol{f}}$

δ_j: j-th element of the vector $\boldsymbol{\delta}$

$\boldsymbol{a} \in \mathbb{R}^{m\times 1}$

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

$\boldsymbol{f} \in \mathbb{R}^{n\times 1}$

$g(\boldsymbol{f})$: $\mathbb{R}^{n\times 1} \to \mathbb{R}$

$$\frac{\partial g(\boldsymbol{f})}{\partial W_{ij}} = \frac{\partial \boldsymbol{f}(W)}{\partial W_{ij}} \frac{\partial g(\boldsymbol{f})}{\partial \boldsymbol{f}}$$

f_k: k-th element of vector $\boldsymbol{f}$

$$= [0 \quad \cdots \quad 0 \quad a_i \quad 0 \quad \ldots \quad 0]\boldsymbol{\delta}$$

$$= a_i \delta_j$$

- Proof (contd):

$$\frac{\partial g(\boldsymbol{f})}{\partial W_{ij}} = a_i \delta_j$$

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

$$\boldsymbol{f} = W^T \boldsymbol{a}$$

$$\Longrightarrow \frac{\partial g(\boldsymbol{f})}{\partial W} = \begin{bmatrix} a_1\delta_1 & \cdots & a_1\delta_n \\ \vdots & \ddots & \vdots \\ a_m\delta_1 & \cdots & a_m\delta_n \end{bmatrix}$$

$$= \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix} \begin{bmatrix} \delta_1 & \cdots & \delta_n \end{bmatrix}$$

$$= \boldsymbol{a}\boldsymbol{\delta}^T$$

$$\Longrightarrow \boxed{\frac{\partial g(\boldsymbol{f})}{\partial W} = \boldsymbol{a}\left(\frac{\partial g(\boldsymbol{f})}{\partial \boldsymbol{f}}\right)^T}$$