$$\boxed{Y = WX \Rightarrow \frac{\partial Y}{\partial W}, \frac{\partial Y}{\partial X}}$$

1. $v(x) = k(x)u(x)$, $k \in R, x \in R^n$, $u, v \in R^m$

$$\Rightarrow \frac{\partial v}{\partial x} = \frac{\partial v}{\partial k}\frac{\partial k}{\partial x} + \frac{\partial v}{\partial u}\frac{\partial u}{\partial x}$$

$$= uk' + (kI)u'$$

$$= uk' + ku'$$

2. $k(x) = \langle u(x), v(x) \rangle$

$$= u^T v$$

$k \in R$, $u, v \in R^m$, $x \in R^n$

$$\Rightarrow \frac{\partial k}{\partial x} = \frac{\partial k}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial k}{\partial v}\frac{\partial v}{\partial x}$$

$$= v^T u' + u^T v'$$

3. $v(x) = A \cdot u(x)$, $u, v \in R^m$, $x \in R^n$, $A \in R^{m \times m}$

$$\Rightarrow \frac{\partial v}{\partial x} = \frac{\partial v}{\partial u}\frac{\partial u}{\partial x} = Au'$$

4. $f_{e_i^T}(A) := e_i^T(A) := (0, \cdots, 0, \underbrace{A}_{i\text{-th pos}}, 0, \cdots, 0)$

here A is any object.

$\therefore f_{e_i}(1) = e_i$, $f_{e_i^T}(1) = e_i^T$

5. $y = Wx$, $W \in R^{m \times n}$

$$\Rightarrow \frac{\partial y}{\partial \tilde{W}} = \frac{\partial y}{\partial W.\mathrm{vec}} = \begin{pmatrix} x^T & & \\ & x^T & \\ & & x^T \end{pmatrix}_{m \times m}$$

$$= \mathrm{diag}(x^T, x^T, x^T)$$

proof: $y_i = w_i^T x$, here $W = \begin{pmatrix} \vdots \\ w_i^T \\ \vdots \end{pmatrix}$

$$\therefore \frac{\partial y_i}{\partial w_j} = \begin{cases} x^T, & i = j \\ 0, & \text{otherwise} \end{cases}$$

$$\therefore \frac{\partial y_i}{\partial \tilde{W}} = f_{e_i^T}(x^T)$$

$$\therefore \frac{\partial Y_j}{\partial \tilde{W}} = \begin{pmatrix} X_j^T & & \\ & \ddots & \\ & & X_j^T \end{pmatrix}_{m\times m} \in R^{m\times(m\cdot n)}$$

$$\therefore \frac{\partial Y_j}{\partial \tilde{W}} = \begin{pmatrix} f_{e_1^T}(X_j^T) \\ \cdots \\ f_{e_m^T}(X_j^T) \end{pmatrix}$$

$$\therefore \frac{\partial Y_{ij}}{\partial \tilde{W}} = f_{e_i^T}(X_j^T)$$

$$\therefore \frac{\partial \tilde{Y}}{\partial \tilde{W}} = \begin{pmatrix} f_{e_1^T}(X_1^T) \\ f_{e_1^T}(X_2^T) \\ \cdots \\ f_{e_1^T}(X_b^T) \\ \cdots \\ f_{e_m^T}(X_1^T) \\ f_{e_m^T}(X_2^T) \\ \cdots \\ f_{e_m^T}(X_b^T) \end{pmatrix} = \begin{pmatrix} X_1^T & & \\ X_2^T & & \\ X_b^T & & \\ & X_1^T & \\ & X_2^T & \\ & X_b^T & \end{pmatrix} = \begin{pmatrix} X^T & & \\ & \cdots & \\ & & X^T \end{pmatrix}_{m\times m}$$

$$\therefore \frac{\partial Y}{\partial \tilde{W}} = \begin{pmatrix} f_{e_1^T}(x^T) \\ \cdots \\ f_{e_m^T}(x^T) \end{pmatrix}$$

$$= \begin{pmatrix} x^T & & \\ & x^T & \\ & & x^T \end{pmatrix}_{m\times m}$$

$$\in R^{m\times(m\cdot n)}$$

6. $Y = WX$, $X \in R^{n\times b}$, $Y \in R^{m\times b}$

$$\Rightarrow \frac{\partial \tilde{Y}}{\partial \tilde{W}} = \frac{\partial Y_{.rvec}}{\partial W_{.rvec}} = \frac{\partial Y_{.vec}}{\partial W_{.vec}}$$

$$= \begin{pmatrix} X^T & & \\ & X^T & \\ & & X^T \end{pmatrix}_{m\times m}$$

$$\in R^{(m\cdot b)\times(m\cdot n)}$$

prof: let $Y = (Y_1, Y_2, Y_b)$, $X = (X_1, X_2, X_b)$

$\because$ $Y_j = WX_j$

8. $Y = WX \Rightarrow \frac{\partial Y}{\partial X}$, here $Y \in R^{m \times b}$, $X \in R^{n \times b}$

solve: $\because Y^T = X^T W^T$

$\therefore \frac{\partial Y^T.vec}{\partial X^T.vec} = \frac{\partial \tilde{Y}^T}{\partial \tilde{X}^T} = \begin{pmatrix} W & & \\ & W & \\ & & W \end{pmatrix}_{b \times b}$

$\in R^{(b \cdot m) \times (b \cdot n)}$

$\therefore \frac{\partial Y^T}{\partial X^T} = \left[\begin{pmatrix} W & & \\ & W & \\ & & W \end{pmatrix}_{b \times b} \right].reshape(b,m,b,n)$

$= A \in R^{b \times m \times b \times n}$

$\because \frac{\partial Y^T_{ij}}{\partial X^T} = \frac{\partial (Y^T)_{ji}}{\partial X^T} = A(j,i,:,:)$

$\therefore \frac{\partial Y}{\partial X^T} = A.swapaxes(1,2)$

$\because \frac{\partial Y}{\partial X_{ij}} = \frac{\partial Y}{\partial (X^T)_{ji}} = \left(\frac{\partial Y}{\partial X^T}\right)(:,:,j,i)$

$\therefore \frac{\partial Y}{\partial X} = \left(\frac{\partial Y}{\partial X^T}\right).swapaxes(3,4)$

$= A.swap\_axes(1,2).swap\_axes(3,4)$

$= A.permute\_axes(2,1,4,3)$

7. $y^T = w^T X$, $X \in R^{n \times b}$, $y \in R^b$

$\Rightarrow \frac{\partial y}{\partial \tilde{X}} = \begin{pmatrix} vec(w,0,0) \\ vec(0,w,0) \\ vec(0,0,w) \end{pmatrix}_b \in R^{b \times (n \cdot b)}$

proof: $\because y_i = w^T X_i$,

here $X = (:, X_i, :)_b$

$y^T = (:, y_i, :)_b$

$\therefore \frac{\partial y_i}{\partial X_i} = w^T$

$\therefore \frac{\partial y_i}{\partial X} = (0, \underset{i}{w}, 0)$,

$\forall i \in \{1, \cdots, b\}$

$\therefore \frac{\partial y}{\partial \tilde{X}} = \begin{pmatrix} \frac{\partial y_1}{\partial \tilde{X}} \\ \vdots \\ \frac{\partial y_b}{\partial \tilde{X}} \end{pmatrix} = \begin{pmatrix} vec(w,0,0) \\ vec(0,w,0) \\ vec(0,0,w) \end{pmatrix}$

10. $z = f(Y)$, $Y = WX$, $z \in R$, $Y \in R^{m \times b}$, $X \in R^{n \times b}$

$\Rightarrow \frac{\partial z}{\partial W} = f'(Y) X^T$

proof: ① method 1:

let $Y = \begin{pmatrix} \vdots \\ y_i^T \\ \vdots \end{pmatrix}$, $X = \begin{pmatrix} \vdots \\ x_j^T \\ \vdots \end{pmatrix}$

$\because y_i = X^T w_i = (x_1, \cdots, x_n) \begin{pmatrix} w_{i1} \\ \vdots \\ w_{in} \end{pmatrix} = \sum_{j=1}^{n} w_{ij} x_j$

$\therefore \frac{\partial y_i}{\partial w_{ij}} = x_j$

$\therefore \frac{\partial z}{\partial w_{ij}} = \frac{\partial z}{\partial y_i} \frac{\partial y_i}{\partial w_{ij}} = \frac{\partial z}{\partial y_i} x_j$

$\therefore \frac{\partial z}{\partial W} := \left(\frac{\partial z}{\partial W}\right).as\_mat = \frac{\partial z}{\partial Y} X^T$

② method 2:

let $Y = \begin{pmatrix} \vdots \\ y_i^T \\ \vdots \end{pmatrix}_m$

9. $z = f(y)$, $y = Wx$, $x \in R^n$, $y \in R^m$, $z \in R$

$\Rightarrow \frac{\partial z}{\partial \tilde{W}} = \frac{\partial z}{\partial y} \frac{\partial y}{\partial \tilde{W}}$

$= \left(\frac{\partial z}{\partial y_1}, \frac{\partial z}{\partial y_m}\right)_{1 \times m} \begin{pmatrix} x^T & \\ & x^T \end{pmatrix}_{m \times m}$

$= \left(\frac{\partial z}{\partial y_1} x^T, \cdots, \frac{\partial z}{\partial y_m} x^T\right)_{1 \times m} \in R^{1 \times (m \cdot n)}$

$\therefore \frac{\partial z}{\partial W} = f'(y)^T x^T$, here $f'(y) \in R^{b \times m}$

(see derivatives 26)

9.3 $\frac{\partial z}{\partial x} \in R^{1 \times n}$

$\frac{\partial z}{\partial W} \in R^{1 \times (m \cdot n)}$

$\because \frac{\partial z}{\partial W}$ not used in chain rule

$\therefore$ we define $\frac{\partial z}{\partial W} := \left(\frac{\partial z}{\partial W}\right).as\_mat$
$\in R^{m \times n}$

if W is not vector.

$\because \frac{\partial z}{\partial \tilde{W}} = \frac{\partial z}{\partial \tilde{Y}} \frac{\partial \tilde{Y}}{\partial \tilde{W}}$

$= \left(\frac{\partial z}{\partial y_1}, \frac{\partial z}{\partial y_m}\right)_{1\times m} \begin{pmatrix} X^T & & \\ & \cdots & \\ & & X^T \end{pmatrix}_{m\times m}$

$= \left(\frac{\partial z}{\partial y_1} X^T, \cdots, \frac{\partial z}{\partial y_m} X^T\right)_m$

$\in R^{1\times(m\cdot n)}$

$\therefore \frac{\partial z}{\partial W} := \left(\frac{\partial z}{\partial W}\right).as\_mat$

$= \begin{pmatrix} \frac{\partial z}{\partial y_1} X^T \\ \frac{\partial z}{\partial y_m} X^T \end{pmatrix} = \frac{\partial z}{\partial Y} X^T$

(here $\frac{\partial z}{\partial Y} = f'(Y) := \left(\frac{\partial z}{\partial Y}\right).as\_mat$)

③ method 3:

$\because y_i^T = w_i^T X$

$\therefore y_i = X^T w_i \quad \therefore \frac{\partial y_i}{\partial w_i} = X^T$

$\therefore \frac{\partial z}{\partial w_i} = \frac{\partial z}{\partial y_i} \frac{\partial y_i}{\partial w_i} = \frac{\partial z}{\partial y_i} X^T$

$\therefore \frac{\partial z}{\partial W} := \left(\frac{\partial z}{\partial W}\right).as\_mat = \frac{\partial z}{\partial Y} X^T$

11. $z = f(Y)$, $Y = WX$, $z \in R$, $Y \in R^{m\times b}$, $X \in R^{n\times b}$

$\Rightarrow \frac{\partial z}{\partial X}$

Solve: $\because Y^T = X^T W^T$

$\therefore \frac{\partial z}{\partial X^T} = \frac{\partial z}{\partial Y^T} W$

$\therefore \frac{\partial z}{\partial X} := \left(\frac{\partial z}{\partial X}\right).as\_mat$

$= W^T \frac{\partial z}{\partial Y}$

$= W^T f'(Y)$