$\boxed{\text{bilinear}(X, Y, A, b)}$

1. $\frac{dL}{dX} := \left(\frac{\partial L}{\partial X}\right)$.reshape(X.shape)

$\forall$ tensor X, here $L \in R$

$dX := \frac{dL}{dX}$, $\forall$ tensor X.

$\therefore \frac{dL}{dX} = \left(\frac{\partial L}{\partial X}\right)$.as_mat, when X is mat.

$\frac{dL}{dX} = \left(\frac{\partial L}{\partial X}\right)$.as_vec, when X is vector.

$\frac{dL}{dX} = \frac{\partial L}{\partial X}$, if X is row vector.

2. $L = L(Z)$, $Z = Y + b$, $L \in R$,

$Y = WX \Rightarrow \frac{dL}{dX}, \frac{dL}{dW}, \frac{dL}{db}$ (see 01018)

solve: $\because \frac{\partial L}{\partial Y} = \frac{\partial L}{\partial Z}$

$\therefore \frac{dL}{dX} := \left(\frac{\partial L}{\partial X}\right)\text{.as\_mat} = W^T \left(\frac{\partial L}{\partial Y}\right)\text{.as\_mat} = W^T \frac{dL}{dZ}$

$\frac{dL}{dW} := \left(\frac{\partial L}{\partial W}\right)\text{.as\_mat} = \left(\frac{\partial L}{\partial Y}\right)\text{.as\_mat}\, X^T = \frac{dL}{dZ} X^T$

$\because Z = Y + B$, $B = (b, b, \cdots)$

$$\therefore \frac{dL}{db} := \left(\frac{\partial L}{\partial b}\right)\text{.as\_vec} = \left(\frac{\partial L}{\partial b}\right)^T$$

$$= \text{sum}\left(\left(\frac{\partial L}{\partial B}\right)\text{.as\_mat},\ \text{dim}2\right)$$

$$= \text{sum}\left(\frac{dL}{dZ},\ \text{dim}2\right)$$

$$= \frac{dL}{dZ}\text{.sum}(2)$$

$$= \frac{dL}{dZ}\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}_b$$

3. linear: $Y = f(X, A, b)$

$y = Ax + b$, $A \in R^{m \times n}$, $b \in R^m$,

$\therefore dX = dY \cdot A$

$dA = (dY)^T X$

$db = (dY)^T \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}_N$, $N = X.\mathrm{dim}(1)$

4. $y = Ax$, $z = f(y)$, $z \in R$, $A \in R^{m \times n}$

$\Rightarrow \frac{dz}{dA} = \frac{dz}{dy} x^T$

5. $z = f(x, y, A) = x^T A y$, $A \in R^{m \times n}$, $z \in R$

$\Rightarrow \frac{dz}{dx}, \frac{dz}{dy}, \frac{dz}{dA}$

solve: $\because \frac{\partial z}{\partial x} = y^T A^T$ $\therefore \frac{dz}{dx} = \left(\frac{\partial z}{\partial x}\right)^T = Ay$

$\because \frac{\partial z}{\partial y} = x^T A$, $\therefore \frac{dz}{dy} = \left(\frac{\partial z}{\partial y}\right)^T = A^T x$

let $v = Ay$, $\therefore \frac{\partial z}{\partial v} = x^T$ $\therefore \frac{dz}{dA} = \frac{dz}{dv} y^T = x y^T$

$X = \begin{pmatrix} \vdots \\ x^T \\ \vdots \end{pmatrix}$, $Y = \begin{pmatrix} \vdots \\ y^T \\ \vdots \end{pmatrix}$

3.3 forward$(X, A, b) \Rightarrow Y$

solve: $\because y = Ax + b$

$\therefore Y = XA^T + b^T$

3.4 backward$(dY) \Rightarrow dX, dW, db$

solve: $\because Y = XA^T + b^T$

$\therefore Y^T = AX^T + b$

$\therefore \frac{dL}{dX^T} = A^T \frac{dL}{dY^T}$ $\therefore \frac{dL}{dX} = \frac{dL}{dY} A$

$\frac{dL}{dA} = \frac{dL}{dY^T} X = \left(\frac{dL}{dY}\right)^T X$

$\frac{dL}{db} = \left(\frac{\partial L}{\partial b}\right)^T = \mathrm{sum}\left(\frac{dL}{dY^T}, \mathrm{dim}2\right) = \left(\frac{dL}{dY}\right)^T \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$

here $\cdot_e$ means elementwise multiplication.

①': $x, y \in R^n$, $x \cdot y := x \cdot_e y \in R^n$

② $x \in R^{1\times n}$, $Y \in R^{m\times n}$

$$x \cdot Y = (x_1, x_n) \cdot (y_1, y_n) := (x_1 y_1, x_n y_n)_{m\times n}$$
$$= x \cdot_e Y$$

②': $x \in R^m$, $Y \in R^{m\times n}$,

$$\Rightarrow x \cdot Y := x \cdot_e Y \in R^{m\times n}$$

③ $X \in R^{m\times n}$, $Y \in R^{m\times n}$

$$X \cdot Y = (X_1, X_n) \cdot (Y_1, Y_n) := (X_1^T Y_1, X_n^T Y_n)_{1\times n}$$
$$= (X \cdot_e Y).sum(1, keepdim)$$

③' $X, Y \in R^{m\times n}$,

$$X \cdot Y = \begin{pmatrix} x_1^T \\ x_m^T \end{pmatrix} \cdot \begin{pmatrix} y_1^T \\ y_m^T \end{pmatrix} := \begin{pmatrix} x_1^T y_1 \\ x_m^T y_m \end{pmatrix}_m = (X \cdot_e Y).sum(2)$$

6. $Z = X^T A Y$, $A \in R^{m\times n}$,

$$X = (x_1, x_2, x_{bx}) \in R^{m\times bx}$$
$$Y = (y_1, y_2, y_{by}) \in R^{n\times by}$$

6.3 ∵ $Z = \begin{pmatrix} x_1^T \\ x_2^T \\ x_{bx}^T \end{pmatrix} A (y_1, y_2, y_{by})$

$$= \begin{pmatrix} x_1^T \\ x_2^T \\ x_{bx}^T \end{pmatrix} (Ay_1, Ay_2, Ay_{by})$$

$$= \begin{pmatrix} x_1^T A y_1 & x_1^T A y_2, & x_1^T A y_{by} \\ x_2^T A y_1, & & \\ x_{bx}^T A y_1, & \cdots & \end{pmatrix}_{bx \times by}$$

7. definition

① $x, y \in R^{1\times n}$,

$$x \cdot y = (x_1, x_n) \cdot (y_1, y_n) := (x_1 y_1, x_n y_n)_{1\times n} = x \cdot_e y$$

$\therefore \frac{dL}{dX} = \left(\frac{dL}{dz_1} Ay_1, \frac{dL}{dz_2} Ay_2, \cdots\right)$

$= \left(\frac{dL}{dz_1}, \frac{dL}{dz_2}, \cdots\right) \bullet (Ay_1, Ay_2, \cdots)$

$= \frac{dL}{dz} \cdot AY = \frac{dL}{dz} \dot{e} AY$

(· is defined in 7.②)

② $\because z = X \cdot AY = Y \cdot A^T X$ (· in 7.③)

$\therefore \frac{dL}{dY} = \frac{dL}{dz} \cdot A^T X$ (· in 7.②)

③ $\frac{dL}{dA} = \sum_{i=1}^{b} \frac{dL}{dz_i} \frac{dz_i}{dA} = \sum^{b} \left(\frac{dL}{dz_i} x_i\right) y_i^T$

$= \left(\frac{dL}{dz} \cdot X\right) Y^T$ (· in 7.②)

9. definition

8. $z = f(X, Y, A) = X \cdot AY$

$= X \cdot (AY) = (x_1, x_2, x_b) \cdot (Ay_1, Ay_2, Ay_b)$

$:= (x_1^T Ay_1, x_2^T Ay_2, x_b^T Ay_b)$

(· is defined in 7.③)

$A \in R^{m \times n}$, $X \in R^{m \times b}$, $Y \in R^{n \times b}$

$z \in R^{1 \times b}$, $\frac{dL}{dz} \in R^{1 \times b}$ is given.

$\Rightarrow \frac{dL}{dX}, \frac{dL}{dY}, \frac{dL}{dA}$

solve: let $z = (z_1, z_2, \cdots, z_b)$

① $\because \frac{\partial L}{\partial x_i} = \frac{\partial L}{\partial z_i} \frac{\partial z_i}{\partial x_i} = \frac{\partial L}{\partial z_i} y_i^T A^T$

$\therefore \frac{dL}{dx_i} = \frac{dL}{dz_i} Ay_i$

① $x, y \in R^n$, $x \cdot y := \langle x, y \rangle$
$= \sum_{i=1}^{n} x_i y_i \in R$

② $x \in R^L$, $Y \in R^{L \times m \times n}$

$x \cdot Y := \sum_{i=1}^{L} x_i Y_i = (X \cdot_e Y).\mathrm{sum}(1)$
$\in R^{m \times n}$

here $Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_L \end{pmatrix}$

10. $z = f(x, y, A)$, $L = L(z)$, $L \in R$,

$z = \begin{pmatrix} z_1 \\ \vdots \\ z_L \end{pmatrix}$, $A = \begin{pmatrix} A_1 \\ \vdots \\ A_L \end{pmatrix}$

$z_i = x^T A_i y$, $A \in R^{L \times m \times n}$,
$z \in R^L$, $x \in R^m$, $y \in R^n$

given $\frac{dL}{dz} \Rightarrow \frac{dL}{dx}, \frac{dL}{dy}, \frac{dL}{dA}$

solve: ① $\frac{dL}{dx} = \sum_{i=1}^{L} \frac{dL}{dz_i} \frac{dz_i}{dx} = \sum_{i=1}^{L} \frac{dL}{dz_i} A_i y$

$= \left(\sum_{i=1}^{L} \frac{dL}{dz_i} A_i \right) y$

$= \left(\frac{dL}{dz} \cdot A \right) y$, ($\cdot$ in 9, ②)

② $\frac{dL}{dy} = \sum_{i=1}^{L} \frac{dL}{dz_i} A_i^T x = \left(\sum_{i=1}^{L} \frac{dL}{dz_i} A_i \right)^T x$

$= \left(\frac{dL}{dz} \cdot A \right)^T x$

③ $\frac{dL}{dA_i} = \frac{dL}{dz_i} \frac{dz_i}{dA_i} = \frac{dL}{dz_i} x y^T$

$\therefore \frac{dL}{dA} = \begin{pmatrix} \frac{dL}{dA_1} \\ \vdots \end{pmatrix} = \begin{pmatrix} \cdots \\ \frac{dL}{dz_i} x y^T \\ \cdots \end{pmatrix}$

11. $Z = f(X, Y, A)$, $z_{ij} = X_j^T A_i Y_j$

$A = \begin{pmatrix} A_1 \\ A_L \end{pmatrix} \in R^{L \times m \times n}$, $X = (X_1 X_b) \in R^{m \times b}$, $Y = (Y_1 Y_b) \in R^{n \times b}$,

$Z = (Z_1 Z_b) = \begin{pmatrix} z_1^T \\ z_L^T \end{pmatrix} \in R^{L \times b} \Rightarrow \frac{dL}{dX}, \frac{dL}{dY}, \frac{dL}{dA}$.

11.4. method 2 (see 10)

$\because \frac{dL}{dZ} \Rightarrow \frac{dL}{dZ_j}$

$\therefore$ ① $\frac{dL}{dX_j} = \left(\frac{dL}{dZ_j} \cdot A\right) Y_j$ ($\cdot$ in 9.②) $\Rightarrow \frac{dL}{dX}$

② $\frac{dL}{dY_j} = \left(\frac{dL}{dZ_j} \cdot A\right)^T X_j \Rightarrow \frac{dL}{dY}$

③ $\frac{dL}{dA} = \sum_{j=1}^{b} \begin{pmatrix} \frac{dL}{dz_{1j}} X_j Y_j^T \\ \vdots \\ \frac{dL}{dz_{lj}} X_j Y_j^T \end{pmatrix}$

$\therefore \frac{dL}{dA_i} = \sum_{j=1}^{b} \frac{dL}{dz_{ij}} X_j Y_j^T$

$= \left(\frac{dL}{dz_i^T} \cdot X\right) Y^T$ ($\cdot$ in 7.②)

$= \left(\frac{dL}{dz_i^T} \cdot_c X\right) Y^T$

$\Rightarrow \frac{dL}{dA}$

11.2. $\because z_{ij} = X_j^T A_i Y_j$

$\therefore z_i^T = f_i(X, Y, A_i)$,

(f_i defined in 8).

$Z_j = F_j(X_j, Y_j, A)$,

(F_j defined in 10)

11.3 method 1. (see 8)

$\because \frac{dL}{dZ} \Rightarrow \frac{dL}{dz_i^T}$

$\therefore$ ① $\frac{dL}{dX} = \sum_{i=1}^{l} \left(\frac{dL}{dz_i^T} \cdot A_i Y\right)$ ($\cdot$ in 7.②)

② $\frac{dL}{dY} = \sum_{i=1}^{l} \left(\frac{dL}{dz_i^T} \cdot A_i^T X\right)$ ($\cdot$ in 7.②)

③ $\frac{dL}{dA_i} = \left(\frac{dL}{dz_i^T} \cdot X\right) Y^T$ ($\cdot$ in 7.②)

$\Rightarrow \frac{dL}{dA}$

12. bilinear

$$Z = f(X, Y, A, b).$$

$$z_{ij} = x_i^T A_j y_i + b_j$$

$$X = \begin{pmatrix} x_1^T \\ x_b^T \end{pmatrix}, \quad Y = \begin{pmatrix} y_1^T \\ y_b^T \end{pmatrix}, \quad Z = \begin{pmatrix} z_1^T \\ z_b^T \end{pmatrix}$$

$$A = \begin{pmatrix} A_1 \\ A_L \end{pmatrix}, \quad X \in R^{b\times m}, \quad Y \in R^{b\times n}$$

$$Z \in R^{b\times L}, \quad A \in R^{L\times m\times n}, \quad b \in R^L$$

12.3 forward$(X, Y, A, b) \Rightarrow Z$

solve: let $Z = (Z_1, \cdots, Z_L)$

$$\because Z_j = \begin{pmatrix} x_1^T A_j y_1 \\ \cdots \\ x_b^T A_j y_b \end{pmatrix} = XA_j \cdot Y, \quad (\cdot \text{ in } 7.③')$$

$$\therefore Z_j = \left[(XA_j) \cdot_e Y\right].\text{sum}(2) \Rightarrow Z$$

$$Z \leftarrow Z + b^T$$

12.4 backward$\left(\frac{dL}{dZ}\right) \Rightarrow \frac{dL}{dX}, \frac{dL}{dY}, \frac{dL}{dA}, \frac{dL}{db}$

① $\because \frac{dL}{dX^T} = \sum_{i=1}^{L}\left(\frac{dL}{dZ_i^T} \cdot_e A_i Y^T\right)$ (see 11.3.①)

$$\therefore \frac{dL}{dX} = \sum_{i=1}^{L}\left(\frac{dL}{dZ_i} \cdot_e YA_i^T\right) = \sum_{i=1}^{L}\left(\frac{dL}{dZ_i} \cdot_e (YA_i^T)\right)$$

② $\because$ 11.3 ①, ②.

$$\therefore \frac{dL}{dY} = \sum_{i=1}^{L}\left(\frac{dL}{dZ_i} \cdot_e XA_i\right)$$

③ $\frac{dL}{dA_i} = \sum_{j=1}^{b} \frac{dL}{dz_{ji}} x_j y_j^T$ (see 11.4.③)

$$= \left(\frac{dL}{dZ_i} \cdot_e X\right)^T Y$$

$$\therefore \frac{dL}{dA} = \begin{pmatrix} \frac{dL}{dA_1} \\ \cdots \\ \frac{dL}{dA_L} \end{pmatrix} \in R^{L\times m\times n}$$

④ $\frac{dL}{db} = \left(\frac{dL}{dZ}\right).\text{sum}(1)$