Gradient of the MatMul operation

$$Y = XW \qquad \frac{\partial \phi}{\partial Y} \qquad \frac{\partial \phi}{\partial X} \qquad \frac{\partial \phi}{\partial W}$$

$$\frac{\partial \phi}{\partial X}$$

$X = [N, D]$ $\quad N = 1$, $D = 3$

$W = [D, M]$ $\quad M = 4$

$Y = [N, M]$

$$Y = \underset{(1,3)}{\begin{bmatrix} x_{11} & x_{12} & x_{13} \end{bmatrix}} \underset{(3,4)}{\begin{bmatrix} w_{11} & w_{12} & w_{13} & w_{14} \\ w_{21} & w_{22} & w_{23} & w_{24} \\ w_{31} & w_{32} & w_{33} & w_{34} \end{bmatrix}} =$$

$$\begin{bmatrix} (x_{11}w_{11} + x_{12}w_{21} + x_{13}w_{31}) & (x_{11}w_{12} + x_{12}w_{22} + x_{13}w_{32}) & (x_{11}w_{13} + x_{12}w_{23} + x_{13}w_{33}) & (x_{11}w_{14} + x_{12}w_{24} + x_{13}w_{34}) \end{bmatrix}$$

$$\frac{\partial \phi}{\partial y} = \begin{bmatrix} dy_{11} & dy_{12} & dy_{13} & dy_{14} \end{bmatrix}$$

$$\frac{\partial \phi}{\partial x} = \frac{\partial \phi}{\partial y} \cdot \frac{\partial y}{\partial x}$$

$$\begin{bmatrix} (x_{11}w_{11} + x_{12}w_{21} + x_{13}w_{31}) & (x_{11}w_{12} + x_{12}w_{22} + x_{13}w_{32}) & (x_{11}w_{13} + x_{12}w_{23} + x_{13}w_{33}) & (x_{11}w_{14} + x_{12}w_{24} + x_{13}w_{34}) \end{bmatrix}$$

$$\frac{\partial y}{\partial x} = \begin{bmatrix} w_{11} & w_{21} & w_{31} \\ w_{12} & w_{22} & w_{32} \\ w_{13} & w_{23} & w_{33} \\ w_{14} & w_{24} & w_{34} \end{bmatrix} = W^T$$

$$\frac{\partial \phi}{\partial x} = \frac{\partial \phi}{\partial y} \cdot \frac{\partial y}{\partial x} = \underset{[N,M]}{\frac{\partial \phi}{\partial y}} \cdot \underset{[M,D]}{W^T} = [N,D]$$

$$\underset{[D,M]}{\frac{\partial \phi}{\partial W}} = \underset{[D,N]}{X^T} \cdot \underset{[N,M]}{\frac{\partial \phi}{\partial y}}$$