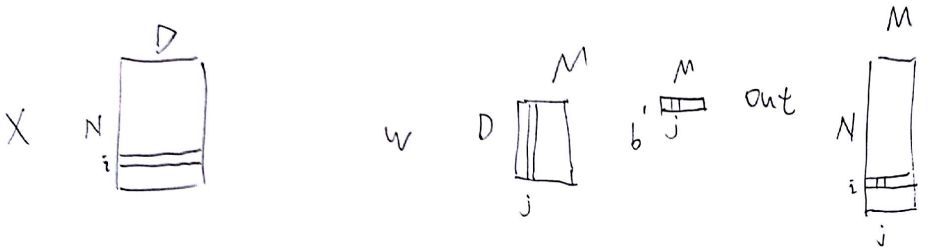


$$out = X w + b \Rightarrow out_{ij} = X_i W_j + b_j$$

$$\frac{\partial\, out_{ij}}{\partial\, X_i} = W_j$$

$$\frac{\partial\, out_{ij}}{\partial\, W_j} = X_i$$

$$\frac{\partial\, out_{ij}}{\partial\, b_j} = 1$$

$\frac{\partial L}{\partial \text{out}}$ given

$N \times M$

$$\frac{\partial L}{\partial X} = \frac{\partial L}{\partial \text{out}} \frac{\partial \text{out}}{\partial X}$$

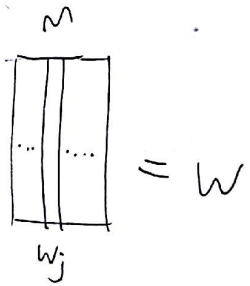
$N \times D$ $\quad N \times M$ $\quad (N \times M) \times (N \times D)$

We know out_i is only computed from X_i, not other X's.

For one data point i

$$\frac{\partial L}{\partial X_i} = \frac{\partial L}{\partial \text{out}_i} \cdot \frac{\partial \text{out}_i}{\partial X_i}$$

D $\quad M$ $\quad M \times D$



To vectorize for all data points,

$$\frac{\partial L}{\partial X} = \frac{\partial L}{\partial \text{out}} \cdot \frac{\partial \text{out}}{\partial X}$$

$N \times D$ $\quad N \times M$ $\quad N \times (M \times D)$

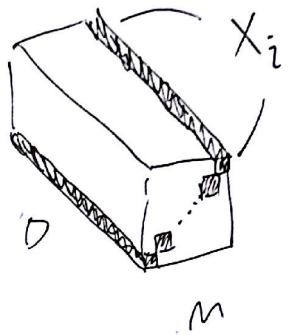
$$\frac{\partial L}{\partial W} = \frac{\partial L}{\partial out} \cdot \frac{\partial out}{\partial W}$$

$D \times M$ | $N \times M$ | $(N \times M) \times D \times M$

For one point i,

$$\frac{\partial L}{\partial W} = \frac{\partial L}{\partial out_i} \cdot \frac{\partial out_i}{\partial W}$$

$D \times M$ | M | $M \times (D \times M)$



$$= \frac{\partial L}{\partial out_i} \times X_i$$

M | D

For all data points,

$$\frac{\partial L}{\partial W} = \frac{\partial L}{\partial out} \times X$$

$N \times D \times M$ | $N \times M$ | $N \times D$

take average

$\frac{\partial L}{\partial W}$

$D \times M$

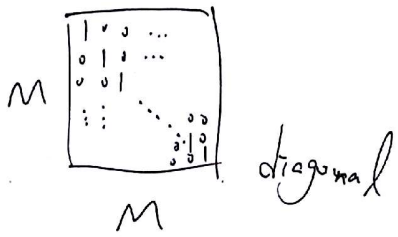
$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial out} \frac{\partial out}{\partial b}$$

$[M]$ $[N \times M]$ $[(N \times M) \times M]$

For one data point i,

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial out_i} \frac{\partial out_i}{\partial b}$$

$[M]$ $[M]$ $[M \times M]$



$$= \frac{\partial L}{\partial out_i}$$

$[M]$

To vectorize for all data points,

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial out}$$

$N \times M$ $N \times M$