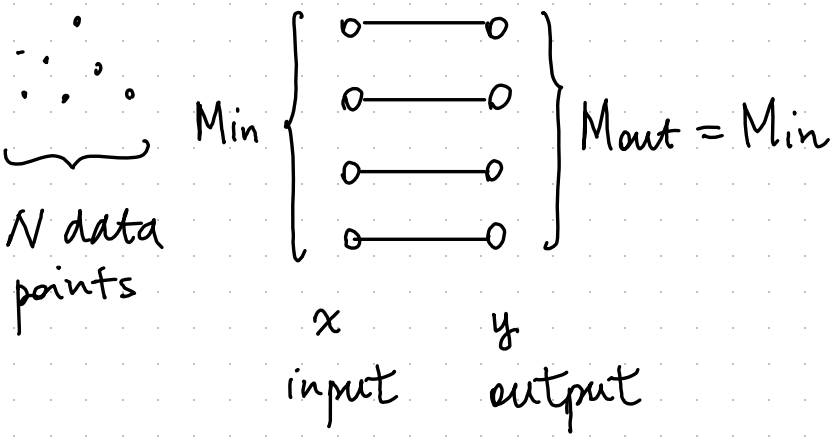
NONLINEAR ACTIVATIONS



$x \in \mathbb{R}^{N \times M_{in}}$

N size of dataset

number of units in the layer

$y \in \mathbb{R}^{N \times M_{out}} = \mathbb{R}^{N \times M_{in}}$

same size of input

$h : \mathbb{R} \longrightarrow \mathbb{R}$

$y = h(x)$ in the sense that

$y_{ij} = h(x_{ij})$ for $i = 0, \dots, N-1$

$j = 0, \dots, M_{in} - 1$

Assuming that we know

$$\frac{\partial L}{\partial y} \in \mathbb{R}^{M_{out} \times N},$$

we compute

$$\frac{\partial L}{\partial x} \in \mathbb{R}^{M_{out} \times N}$$

$$\frac{\partial L}{\partial x_{ij}} = \frac{\partial L}{\partial y_{ij}} h'(x_{ij})$$

Hence,

$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial y} \odot h(x)$$

↑ Hadamard product (component by component)