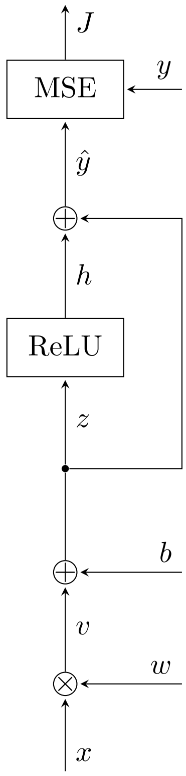
1 Algorithmic Differentiation

Algorithmic differentiation allows to differentiate in a complicated computational graph just by combining gradients of individual graph elements.
This exercise aims to teach the difference between forward and backward differentiation and should show, why the computational graph has to be analyzed before backward differentiation is possible.



(a). Obtain a valid execution order of the operations in the displayed graph. Start with the cost J and determine, which values are necessary to apply the current operation, e.g. $\mathrm{MSE}(\cdot, \cdot)$.

(b). Calculate each value in the graph. Use the seed values $y = 7$, $b = -3$, $w = 2$, and $x = 5$.

(c). Derive the symbolic partial derivatives of each operation independently.

(d). Calculate the numeric forward gradients for each graph element using the following seeds:
  a) $\dot{x}_{\text{seed}} = 1$ others 0
  b) $\dot{w}_{\text{seed}} = 1$ others 0
  c) $\dot{b}_{\text{seed}} = 1$ others 0
  d) $\dot{y}_{\text{seed}} = 1$ others 0

(e). Calculate the numeric backward gradient with the seed $\bar{J}_{\text{seed}} = 1$.

Hint:

$$\text{heaviside}(z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0 \end{cases}$$

Solution

Operation	Value	Symbolic gradients
x	5	
w	2	
b	-3	
y	7	
$v = wx$	10	$\frac{\partial v}{\partial w} = x, \; \frac{\partial v}{\partial x} = w$
$z = v + b$	7	$\frac{\partial z}{\partial v} = 1, \; \frac{\partial z}{\partial b} = 1$
$h = \mathrm{ReLU}(z)$	7	$\frac{\partial h}{\partial z} = \mathrm{heaviside}(z)$
$\hat{y} = h + z$	14	$\frac{\partial y}{\partial h} = 1, \; \frac{\partial y}{\partial z} = 1$
$J = \mathrm{MSE}(\hat{y}, y)$	49	$\frac{\partial J}{\partial \hat{y}} = 2(\hat{y} - y), \; \frac{\partial J}{\partial y} = -2(\hat{y} - y)$

Operation	Forward mode				Backward mode
	$\dot{b}_{\mathrm{seed}} = 1$	$\dot{w}_{\mathrm{seed}} = 1$	$\dot{x}_{\mathrm{seed}} = 1$	$\dot{y}_{\mathrm{seed}} = 1$	$\bar{J}_{\mathrm{seed}} = 1$
x	$\dot{x} = 0$	$\dot{x} = 0$	$\dot{x} = 1$	$\dot{x} = 0$	$\bar{x} = 56$
w	$\dot{w} = 0$	$\dot{w} = 1$	$\dot{w} = 0$	$\dot{w} = 0$	$\bar{w} = 140$
b	$\dot{b} = 1$	$\dot{b} = 0$	$\dot{b} = 0$	$\dot{b} = 0$	$\bar{b} = 28$
y	$\dot{y} = 0$	$\dot{y} = 0$	$\dot{y} = 0$	$\dot{y} = 1$	$\bar{y} = -14$
$v = wx$	$\dot{v} = 0$	$\dot{v} = 5$	$\dot{v} = 2$	$\dot{v} = 0$	$\bar{v} = 28$
$z = v + b$	$\dot{z} = 1$	$\dot{z} = 5$	$\dot{z} = 2$	$\dot{z} = 0$	$\bar{z} = 28$
$h = \mathrm{ReLU}(z)$	$\dot{h} = 1$	$\dot{h} = 5$	$\dot{h} = 2$	$\dot{h} = 0$	$\bar{h} = 14$
$\hat{y} = h + z$	$\dot{y} = 2$	$\dot{y} = 10$	$\dot{y} = 4$	$\dot{y} = 0$	$\bar{v} = 14$
$J = \mathrm{MSE}(\hat{y}, y)$	$\dot{J} = 28$	$\dot{J} = 140$	$\dot{J} = 56$	$\dot{J} = -14$	$\bar{J} = 1$

Example:

$$\begin{aligned} \dot{v} &= \frac{\partial}{\partial v} b = \frac{\partial}{\partial v} w \frac{\partial}{\partial w} b + \frac{\partial}{\partial v} x \frac{\partial}{\partial x} b \\ &= \frac{\partial}{\partial v} w \dot{w} + \frac{\partial}{\partial v} x \dot{x} \end{aligned}$$

$$\bar{v} = \frac{\partial}{\partial J} v = \cdots = \bar{z} \frac{\partial}{\partial z} v$$