Exercice 1. Computational Graph Introduction

1) $\frac{\partial f}{\partial q} = z = -4$

2) $\frac{\partial q}{\partial x} = 1$

3) $\frac{\partial q}{\partial y} = 1$

4) $\frac{\partial f}{\partial z} = x + y = -2 + 5 = 3$

5) $\frac{\partial f}{\partial y} = xz + yz = z = -4$

6) $\frac{\partial f}{\partial x} = xz + yz = z = -4$

Exercice 2: Backpropagation in a computational Graph

$z = w_0 x_0 + w_1 x_1 + w_2$ $\qquad \alpha = \frac{1}{1 + e^{-z}} = \frac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$

1) $\frac{\partial \alpha}{\partial x_0} =$ we define $u(x_0) = 1 + e^{-z}$ with $z = w_0 x_0 + w_1 x_1 + w_2$ with $\alpha(x_0) = \frac{1}{u(x_0)}$

so $\frac{\partial \alpha}{\partial x_0} = \frac{\partial}{\partial x_0} \times \left(\frac{1}{u(x_0)}\right) = -\frac{u'(x_0)}{(u(x_0))^2}$

$u(x_0) = 1 + e^{-(w_0 x_0 + w_1 x_1 + w_2)}$

$u'(x_0) = -w_0 \, e^{-(w_0 x_0 + w_1 x_1 + w_2)} = -w_0 e^{-z}$

$\frac{\partial \alpha}{\partial x_0} = -\frac{(-w_0) e^{-z}}{u(x_0)^2} = \frac{w_0 e^{-z}}{(1 + e^{-z})^2}$ AN

$\boxed{\text{or}}$ derivative of sigmoid fct : $\frac{d(\sigma(x))}{dx} = \sigma(x)(1 - \sigma(x))$

$\Leftrightarrow \frac{\partial \alpha}{\partial z} = \alpha(1 - \alpha)$ and $\frac{\partial z}{\partial x_0} = w_0$

Chain Rule: $\frac{\partial \alpha}{\partial x_0} = \frac{\partial \alpha}{\partial z} \times \frac{\partial z}{\partial x_0}$

Another method

2) $\frac{\partial \alpha}{\partial w_0} = \frac{\partial}{\partial w_0} \times \left(\frac{1}{u(w_0)}\right) = -\frac{u'(w_0)}{(u(w_0))^2}$

$u'(w_0) = -x_0 e^{-z}$

$$\frac{\partial \alpha}{\partial w_0} = \frac{-(-x_0 e^{-z})}{(1+e^{-z})^2} = \frac{x_0 e^{-z}}{(1+e^{-z})^2} \underset{A.N}{=}$$

3) we use chain Rule: $\frac{\partial \alpha}{\partial x_1} = \frac{\partial \alpha}{\partial z} \times \frac{\partial z}{\partial x_1}$

$$\frac{\partial \alpha}{\partial z} = \frac{\partial}{\partial z}\left(\frac{1}{u(z)}\right) = -\frac{u'(z)}{u(z)^2} = \frac{-(-e^{-z})}{(1+e^{-z})^2} = \frac{e^{-z}}{(1+e^{-z})^2}$$

$\frac{\partial z}{\partial x_1} = w_1$

We combine both $\rightarrow$ $\frac{\partial \alpha}{\partial x_1} = w_1 \frac{e^{-z}}{(1+e^{-z})^2} = w_1 \alpha(1-\alpha) \approx -3 \times (0.197) \approx -0.59$

4) $\frac{\partial \alpha}{\partial w_1} = \frac{\partial}{\partial w_1} \times \left(\frac{1}{u(w_1)}\right)$ with $u(w_1) = 1 + e^{-z}$

so $\frac{\partial \alpha}{\partial w_1} = -\frac{u'(w_1)}{(u(w_1))^2} = -\frac{(-x_1 e^{-z})}{(1+e^{-z})^2} = x_1 \frac{e^{-z}}{(1+e^{-z})^2}$

so $\frac{\partial \alpha}{\partial w_1} = x_1 \alpha(1-\alpha) \approx -2 \times 0 \times 0.197 \approx -0.39$

5) $\frac{\partial \alpha}{\partial w_2} = \frac{\partial}{\partial w_2}\left(\frac{1}{u(w_2)}\right)$ with $u(w_2) = 1 + e^{-z}$

$$\frac{\partial \alpha}{\partial w_2} = -\frac{u'(w_2)}{(u(w_2))^2} = -\frac{(1 e^{-z})}{(1+e^{-z})^2} = \frac{-(e^{-z})}{(1+e^{-z})^2} = \frac{e^{-z}}{(1+e^{-z})^2}$$

so $\frac{\partial \alpha}{\partial w_2} = \frac{e^{-z}}{(1+e^{-z})^2} = \alpha(1-\alpha) \approx 0.20 \; (0.197)$

Exercice 3 : Backpropagation focus on dimensions & derivatives

Given Data: $x^{(i)} \in \mathbb{R}^{D_x \times 1}$

$y^{(i)} \in \mathbb{R}^{1 \times 1}$ (scalar)

$z_1 \in \mathbb{R}^{D_{a_1} \times 1}$

1)

$$\underset{\in D_{a_1} \times 1}{z_1} = \underbrace{W_1 \; x^{(i)}}_{\in D_{a_1} \times 1} + \underset{\in D_{a_1} \times 1}{b_1}$$

$x^{(i)} \in D_x \times 1$ so $W_1 \in \mathbb{R}^{D_{a_1} \times D_x}$ so that their multiplication $\in D_{a_1} \times 1$.

$\rightarrow W_1 \in \mathbb{R}^{D_{a_1} \times D_x}$ and $b_1 \in D_{a_1} \times 1$

- $a_1 = ReLU(z_1)$ so $a_1 \in \mathbb{R}^{D_{a_1} \times 1}$

- $\hat{y}^{(i)} = \sigma(z_2)$ so $z_2 \in \mathbb{R}^{1 \times 1}$ since $\hat{y}^{(i)}$ is a scalar

so,

$$\underset{\in \mathbb{R}^{1 \times 1}}{z_2} = \underbrace{W_2 \; a_1}_{\mathbb{R}^{1 \times 1}} + \underset{\in \mathbb{R}^{1 \times 1}}{b_2}$$

So $W_2 \in \mathbb{R}^{1 \times D_{a_1}}$ since $W_2 \cdot a_1 \in \mathbb{R}^{1 \times 1}$

and $b_2 \in \mathbb{R}^{1 \times 1}$

- if we [illegible] multiple example ("m")

  i X will be $\in D_x \times m$

  Y will be $\in 1 \times m$

Biases and weights don't change

2) $\frac{\partial J}{\partial \hat{y}^{(i)}} = \frac{\partial J}{\partial L^{(i)}} \times \frac{\partial L^{(i)}}{\partial \hat{y}^{(i)}}$

* $\frac{\partial J}{\partial L^{(i)}} = -\frac{1}{m}$
* $\frac{\partial L^{(i)}}{\partial \hat{y}^{(i)}} = y^{(i)} \times \frac{1}{\hat{y}^{(i)}} + (1 - y^{(i)}) \times \left(\frac{-1}{1 - \hat{y}^{(i)}}\right)$

$= \frac{y^{(i)}}{\hat{y}^{(i)}} - \frac{(1 - y^{(i)})}{1 - \hat{y}^{(i)}}$

$\frac{\partial J}{\partial \hat{y}^{(i)}} = -\frac{1}{m} \times \left(\frac{y^{(i)}}{\hat{y}^{(i)}} - \frac{(1 - y^{(i)})}{1 - \hat{y}^{(i)}} \right) = \delta_1^{(i)}$

3) $\frac{\partial \hat{y}^{(i)}}{\partial z_2} = \hat{y}^{(i)} (1 - \hat{y}^{(i)})$ (the sigmoid function derivation)

$= \sigma(z_2)(1 - \sigma(z_2)) = \delta_2^{(i)}$

4) $\frac{\partial z_2}{\partial a_1} = W_2 \iff \delta_3^{(i)}$

5) $\frac{\partial a_1}{\partial z_1} = \begin{cases} 1 & \text{if } z_1 > 0 \\ 0 & \text{if } z_1 \leq 0 \end{cases}$ ReLU function $\begin{cases} x \text{ when } x > 0 \\ 0 \text{ when } x \leq 0 \end{cases}$

6) $\frac{\partial z_1}{\partial W_1} = x^{(i)} \iff \delta_5^{(i)}$

7) $\frac{\partial J}{\partial W_1}$: to compute this we will use the chain rule using previous results

$\frac{\partial J}{\partial W_1} = \frac{\partial J}{\partial \hat{y}^{(i)}} \times \frac{\partial \hat{y}^{(i)}}{\partial z_2} \times \frac{\partial z_2}{\partial a_1} \times \frac{\partial a_1}{\partial z_1} \times \frac{\partial z_1}{\partial W_1}$

$\frac{\partial J}{\partial W_1} = \delta_1 \times \delta_2 \times \delta_3 \times \delta_4 \times \delta_5$