Computational Graph Introduction - Mia Hallage

Exercise 1: Simple computational graph

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

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

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

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

5) $\frac{\partial f}{\partial \cancel{z}x} = \frac{\partial (qz)}{\partial x} = \frac{\partial f}{\partial q} \cdot \frac{\partial q}{\partial x} = z \cdot \overset{1}{\cancel{4}} = -4$

6) $\frac{\partial f}{\partial y} = \frac{\partial (qz)}{\partial y} = \frac{\partial f}{\partial q} \cdot \frac{\partial q}{\partial y} = z \cdot 1 = -4$

w/ chain rule

Exercise 2: Backpropagation in a computational graph

$z = w_0 x_0 + w_1 x_1 + w_2$

$\alpha = \sigma(z) = \frac{1}{1+e^{-z}}$

$w_0 = 2$ $\quad w_1 = -3$ $\quad w_2 = -3$ $\quad \frac{\partial L}{\partial \alpha} = 1$

$x_0 = -1$ $\quad x_1 = -2$

$z = 1$ $\quad \alpha \simeq 0,73$

1) $\frac{\partial \alpha}{\partial x_0} = \frac{\partial \alpha}{\partial z} \cdot \frac{\partial z}{\partial x_0}$

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

$g'(u) = -u^{-2}$

$u(z) = 1+e^{-z}$

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

$\sigma'(z) = g'(u(z)) \cdot u'(z) = [-(1+e^{-z})^{-2}] \cdot [-e^{-z}] = \frac{e^{-z}}{(1+e^{-z})^2}$

$\simeq 0,197$

2) $\frac{\partial \alpha}{\partial w_0} = \frac{\partial \alpha}{\partial z} \cdot \frac{\partial z}{\partial w_0} = 0,19\cancel{6}7 \times (-1)$
$= -0,197$

pour les mêmes raisons:

3) $\frac{\partial \alpha}{\partial x_1} = 0,197 \times (-3) = -0,5898$

4) $\frac{\partial \alpha}{\partial w_1} = 0,197 \times (-2) = -0,3932$

5) $\frac{\partial \alpha}{\partial w_2} = \frac{\partial \alpha}{\partial z} \cdot \frac{\partial z}{\partial w_2} = \frac{\partial \alpha}{\partial z} \cdot 1 = 0,197$

Exercise 3: Backpropagation focus on dimensions and derivatives

1) For one example:
w_1 ~~≠~~ : $D_{a1} \times D_x$
b_1 : $D_{a1} \times 1$
w_2 : $1 \times D_{a1}$
b_2 : 1×1 (bcs we want z_2 to be a scalar)

. If we vectorize m examples:
~~xxx~~ Weights do not change shape, so w_1, b_1, w_2, b_2 would stay the same,
However X : $D_x \times m$ (each column is one x^i)
Y ~~5~~ : $1 \times m$

Then Z_1 : $D_{a1} \times m$
A_1 : $D_{a1} \times m$ ($ReLU(Z_1)$)
Z_2 : $1 \times m$
$\hat{Y}$: $1 \times m$ ($\sigma(Z_2)$)

X and Y become matrices

2)
$$J = -\frac{1}{m} \Sigma L^{(i)}$$

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

because

$$\frac{\partial L}{\partial \hat{y}} = \frac{y}{\hat{y}} - \frac{1-y}{1-\hat{y}}$$

as $L^{(i)} = y\log(\hat{y}) + (1-y)\log(1-\hat{y})$

derivate wrt $\hat{y}$ $\downarrow$ $\quad$ $\downarrow$ derivate wrt $\hat{y}$

$$\frac{y}{\hat{y}} \qquad (1-y)\cdot\frac{(-1)}{(1-\hat{y})} = -\frac{(1-y)}{(1-\hat{y})}$$

3) $\hat{y}^{(i)} = \sigma(z_2)$

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

$$\delta_2^{(i)} = \frac{\partial \hat{y}^{(i)}}{\partial z_2} = \hat{y}^{(i)}(1-\hat{y}^{(i)})$$

4) $\cancel{\partial y}$ $\frac{\partial z_2}{\partial a_1} = w_2$

5) $\delta_4^{(i)} = \frac{\partial a_1}{\partial z_1} = \mathrm{ReLU}(z_1^{(i)}) = \begin{cases} 1 & \text{if } z_{1,j} > 0 \\ 0 & \text{if } z_{1,j} \leq 0 \end{cases}$

6) $\delta_5^{(i)} = \frac{\partial z_1}{\partial w_1} = x^{(i)T}$

7)
$$\frac{\partial J}{\partial w_1} = \frac{\partial J}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_2} \cdot \frac{\partial z_2}{\partial a_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_1}$$
$$= \delta_1^{(i)} \cdot \delta_2^{(i)} \cdot \delta_3^{(i)} \cdot \delta_4^{(i)} \cdot (\delta_5^{(i)})^T$$