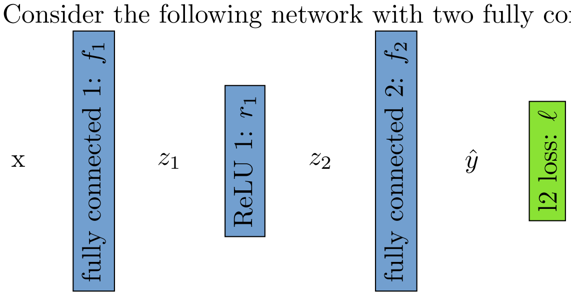
Exercise 2: Forward and back-propataion - Solution

Name: .... UTID:................

In this exercise you'll do some forward and back-prop by hand.
Consider the following network with two fully connected and one ReLU layer:



For some data $x \in R^2$ and a continuous label $y \in R$ the network is defined as follows:

$$z_1 = f_1(x) = \begin{bmatrix} 1 & -2 \\ 0 & 2 \end{bmatrix} x + \begin{bmatrix} 0 \\ -1 \end{bmatrix} \qquad z_2 = r_1(z_1) = \max(z_1, 0)$$

$$\hat{y} = f_2(x) = \begin{bmatrix} 1 & -1 \end{bmatrix} z_2 \qquad \ell(z_2) = \frac{1}{2}\|\hat{y} - y\|^2$$

1. Compute the forward pass and loss of the network for inputs:

a) $x = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, y = 2$

$$\hat{y} = 1, l = \frac{1}{2}$$

b) $x = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, y = 0$

$$\hat{y} = -1, l = \frac{1}{2}$$

c) $x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, y = -2$

$$\hat{y} = -1, l = \frac{1}{2}$$

2. For each input above, compute $\frac{d}{dx}\ell(\hat{y})$ using back-propagation:

a)

$$\begin{aligned}
\frac{d}{d\hat{y}} l(\hat{y}) &= -1 \\
\frac{d\hat{y}}{dz_2} &= \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\
\frac{dz_2}{dz_1} &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \\
\frac{dz_1}{dx} &= \begin{bmatrix} 1 & -2 \\ 0 & 2 \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
\frac{d}{dx} l(\hat{y}) &= \frac{d}{d\hat{y}} l(\hat{y}) \cdot \frac{d\hat{y}}{dz_2} \cdot \frac{dz_2}{dz_1} \cdot \frac{dz_1}{dx} \\
&= \begin{bmatrix} -1 \\ 2 \end{bmatrix}
\end{aligned}$$

b)

$$\begin{aligned}
\frac{d}{d\hat{y}} l(\hat{y}) &= -1 \\
\frac{d\hat{y}}{dz_2} &= \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\
\frac{dz_2}{dz_1} &= \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \\
\frac{dz_1}{dx} &= \begin{bmatrix} 1 & -2 \\ 0 & 2 \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
\frac{d}{dx} l(\hat{y}) &= \frac{d}{d\hat{y}} l(\hat{y}) \cdot \frac{d\hat{y}}{dz_2} \cdot \frac{dz_2}{dz_1} \cdot \frac{dz_1}{dx} \\
&= \begin{bmatrix} 0 \\ 2 \end{bmatrix}
\end{aligned}$$

c)

$$\frac{d}{d\hat{y}}l(\hat{y}) = 1$$

$$\frac{d\hat{y}}{dz_2} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$\frac{dz_2}{dz_1} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\frac{dz_1}{dx} = \begin{bmatrix} 1 & -2 \\ 0 & 2 \end{bmatrix}$$

$$\frac{d}{dx}l(\hat{y}) = \frac{d}{d\hat{y}}l(\hat{y}) \cdot \frac{d\hat{y}}{dz_2} \cdot \frac{dz_2}{dz_1} \cdot \frac{dz_1}{dx} = \begin{bmatrix} 0 \\ -2 \end{bmatrix}$$

3. How many operations (multiplications and additions) do you need to perform in back-propagation. Only count the Jacobean (e.g. $\frac{d}{dz_1}r_1(z_1)$) matrix multiplication operations. How many additional operations would you require to compute $\frac{d}{dW_1}\ell(\hat{y})$ using back-prop (assuming you store all computation from the previous back-prop)?

a) $2+6+6 = 14$; 6, because

$$\frac{d}{W_1}l(\hat{y}) = \frac{d}{dz_1}l(\hat{y}) \cdot \frac{dz_1}{dW_1}$$

b) Same.

c) Same.

4. If you evaluate the same objective using forward propagation (computing $\frac{d}{dx}z_1, \frac{d}{dx}z_2, \frac{d}{dx}z_3, \ldots$ in that order), how many operations would you require? Only count the Jacobean matrix multiplication operations. How many additional operations would you require to compute $\frac{d}{dW_1}\ell(\hat{y})$ using back-prop (assuming you store all computation from the previous back-prop)?

a) $12+6+2 = 20$; $12+6+2 = 20$, because

$$\frac{d}{W_1}l(\hat{y}) = \frac{d}{d\hat{y}}l(\hat{y}) \cdot \frac{d\hat{y}}{dz_2} \cdot \frac{dz_2}{dz_1} \cdot \frac{dz_1}{dW_1}$$

b) Same.

c) Same.