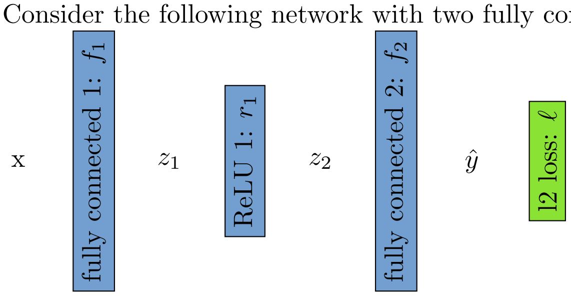
Exercise 2: Forward and back-propataion

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

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

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

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

a)

b)

c)

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)

b)

c)

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)

b)

c)