You have a network with two layers of weights, **W** and **V**, corresponding to the first and second layers, respectively. The non-linear function through which the output of **W** goes is the rectified linear function, $max(0, x)$.

In pseudo-code, the forward pass of the network is as follows:

```
z1 = W x
a1 = max(0, z1)
z2 = V a1
```

The shapes and values of the weight matrices are:

$$\begin{aligned} \mathbf{W} \in \mathbb{R}^{1x1} &\quad [0.20] \\ \mathbf{V} \in \mathbb{R}^{1x1} &\quad [0.40] \end{aligned}$$

Given input $\mathbf{x} \in \mathbb{R}^{1x1} = [0.10]$ and a target of 1, and mean squared error (LMS in our lectures) as the loss function,

- What are the values of z1, a1, z2?

$$\begin{aligned} z_1 &= 0.2 \cdot 0.1 = 0.02 \\ a_1 &= \max(0, 0.2) = 0.02 \\ z_2 &= 0.4 \cdot 0.02 = 0.008 \end{aligned}$$

- What is the value of the error?

$$\text{error} = l = (1 - z_2)^2 = (1 - 0.008)^2 = 0.992^2 = 0.984$$

- What are the gradients of **W** and **V**? (The gradient of $max(0, x)$ is 1 if $x > 0$, 0 otherwise).

$$\begin{aligned} \frac{dl}{dv} &= \frac{dl}{dz_2} \cdot \frac{dz_2}{dv} = -2(1 - z_2)a_1 = -2 \cdot 0.992 \cdot 0.02 = -0.0397 \\ \frac{dl}{da_1} &= \frac{dl}{dz_2} \cdot \frac{dz_2}{da_1} = -2(1 - z_2)v \\ \frac{dl}{dz_1} &= \frac{dl}{a_1} \cdot \frac{da_1}{dz_1} = -2(1 - z_2)v \cdot \mathbf{1}_{z_1 > 0} \\ \frac{dl}{dw} &= \frac{dl}{dz_1} \cdot \frac{dz_1}{dw} = -2(1 - z_2)v \cdot x = -2 \cdot 0.992 \cdot 0.4 \cdot 0.1 = -0.079 \end{aligned}$$