

You have a network with two layers of weights,  $\mathbf{W}$  and  $\mathbf{V}$ , corresponding to the first and second layers, respectively. The non-linear function through which the output of  $\mathbf{W}$  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:

$$\mathbf{W} \in \mathbb{R}^{1 \times 1} \quad [0.20]$$

$$\mathbf{V} \in \mathbb{R}^{1 \times 1} \quad [0.40]$$

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

- What are the values of  $z1$ ,  $a1$ ,  $z2$ ?
- What is the value of the error?
- What are the gradients of  $\mathbf{W}$  and  $\mathbf{V}$ ? (The gradient of  $\max(0, x)$  is 1 if  $x > 0$ , 0 otherwise.