1.

$a^{[1]} = x \in \mathbb{R}^{n_1}$ is a vector.

Let $z^{[l]} = W^{[l]} a^{[l-1]} + b^{[l]}$, which is linear part for a.

Then $a^{[l]} = \sigma(z^{[l]})$ is operating a activation function σ such as sigmoid, ReLU on $z^{[l]}$, which is non-linear part for a.

These operation are to perform linear transformation and non-linear activation on the input x layer by layer until the last layer L, and obtain the output $a^{[L]}$.

So, if we want to get $\nabla_x a^{[L]}$, we can use chain rule to get $\nabla_x a^{[L]} = \frac{\partial a^{[L]}}{\partial a^{[L-1]}} \cdot \frac{\partial a^{[L-1]}}{\partial a^{[L-2]}} \cdot \cdots \cdot \frac{\partial a^{[2]}}{\partial x}$.

Let
- $z^{[l]} = W^{[l]} a^{[l-1]} + b^{[l]}$
- $a^{[l]} = \sigma(z^{[l]})$

Then, $\frac{\partial a^{[l]}}{\partial a^{[l-1]}} = \frac{d\sigma(z^{[l]})}{d z^{[l]}} \cdot W^{[l]}$

Define $\delta^{[l]} := \frac{\partial a^{[L]}}{\partial a^{[l]}} \in \mathbb{R}^{n_l}$

From the output layer:

$\delta^{[L]} = \sigma'(z^{[L]}) = (W^{[L]})^T \delta^{[L]}$

$\delta^{[l]} = (W^{[l+1]})^T \delta^{[l+1]} \odot \sigma'(z^{[l]})$ for $l = L-1, L-2, \cdots, 2$, where $\odot$ is Hadamard product.

Then we well get $\nabla_x a^{[L]} = (W^{[2]})^T \delta^{[2]}$ (We don't need to multiply $\sigma'(z^{[1]})$ since the first layer has no activation)

Algorithm:

```
# n_L = 1 → δ[L] is a scalar
delta = sigma_prime ( z [L] )
for l in (L-1) to 2 :
    delta = ( W[l+1].T * delta ) * sigma_prime ( z [l] )
grad_x = W [2] T * delta
return grad_x
```