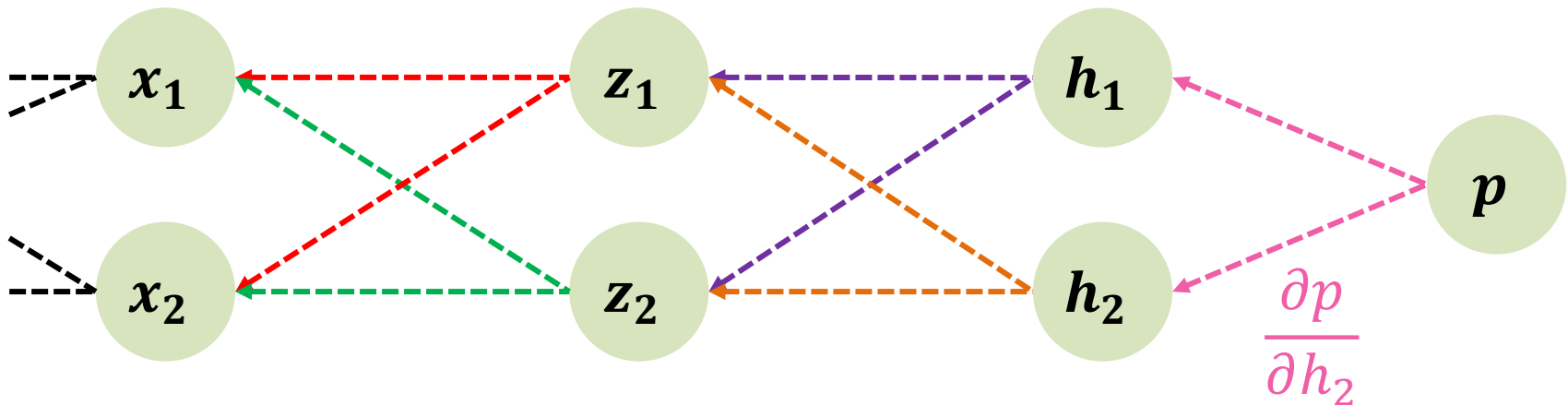# Intro

- In this video we will find out how to apply a chain rule **efficiently**!

Let's look at MLP with 3 hidden layers

Let's drill down to an actual parameter of MLP:

$$h_2 = \sigma(\boldsymbol{w_0} + \boldsymbol{w_1} z_1 + \boldsymbol{w_2} z_2)$$

Gradient Descent:

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial p} \frac{\partial p}{\partial w_1} = \frac{\partial L}{\partial p} \frac{\partial p}{\partial h_2} \frac{\partial h_2}{\partial w_1}$$

$$\frac{\partial p}{\partial h_2}$$

We need to do this efficiently!

We will need these for GD

3: $\quad \dfrac{\partial p}{\partial h_1} \qquad \dfrac{\partial p}{\partial h_2}$

2: $\quad \dfrac{\partial p}{\partial z_1} = \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_1} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_1} \qquad\qquad \dfrac{\partial p}{\partial z_2} = \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_2} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_2}$

1:

$$\dfrac{\partial p}{\partial x_1} = \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_1}\dfrac{\partial z_1}{\partial x_1} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_1}\dfrac{\partial z_1}{\partial x_1} + \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_2}\dfrac{\partial z_2}{\partial x_1} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_2}\dfrac{\partial z_2}{\partial x_1}$$

$$\dfrac{\partial p}{\partial x_2} = \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_1}\dfrac{\partial z_1}{\partial x_2} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_1}\dfrac{\partial z_1}{\partial x_2} + \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_2}\dfrac{\partial z_2}{\partial x_2} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_2}\dfrac{\partial z_2}{\partial x_2}$$

$\dfrac{\partial p}{\partial h_2}$

We can reuse previous computations

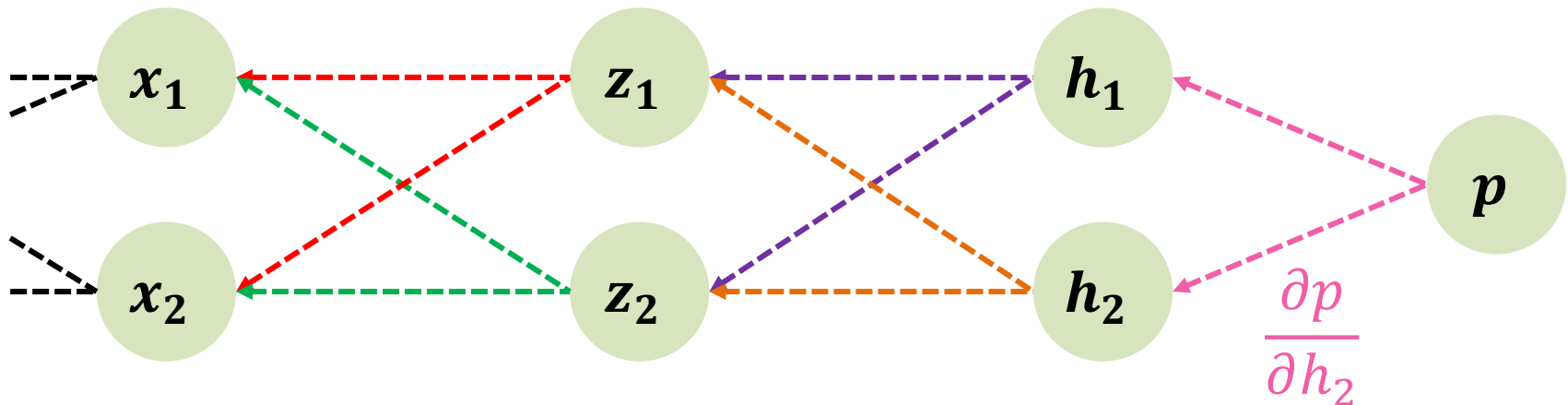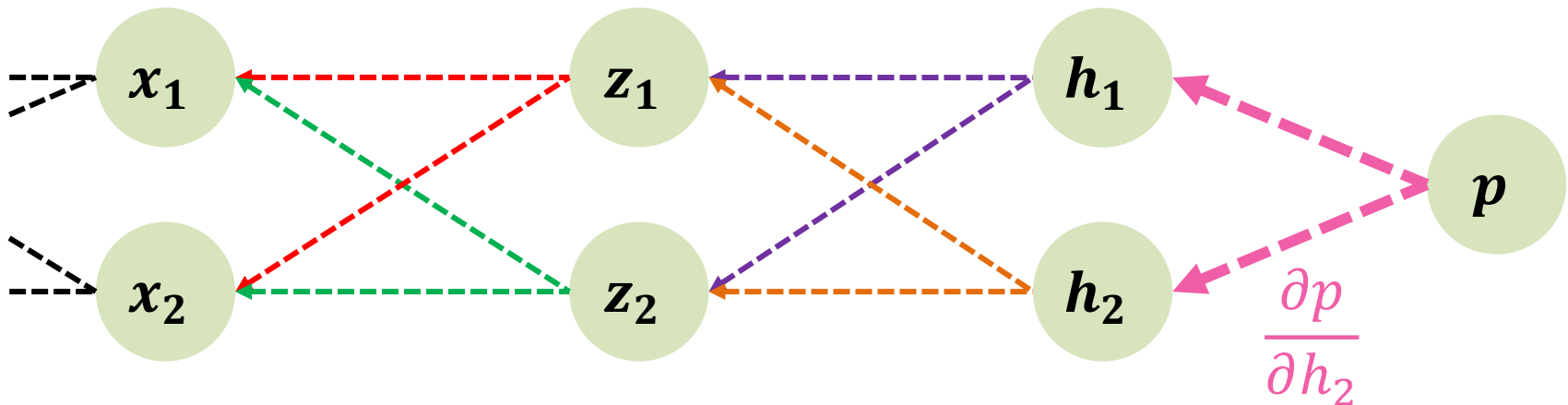3: $\boxed{\dfrac{\partial p}{\partial h_1}}$ $\boxed{\dfrac{\partial p}{\partial h_2}}$ We will need these for GD

2: $\dfrac{\partial p}{\partial z_1} = \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_1} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_1}$ $\qquad$ $\dfrac{\partial p}{\partial z_2} = \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_2} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_2}$

1:

$\dfrac{\partial p}{\partial x_1} = \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_1}\dfrac{\partial z_1}{\partial x_1} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_1}\dfrac{\partial z_1}{\partial x_1} + \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_2}\dfrac{\partial z_2}{\partial x_1} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_2}\dfrac{\partial z_2}{\partial x_1}$

$\dfrac{\partial p}{\partial x_2} = \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_1}\dfrac{\partial z_1}{\partial x_2} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_1}\dfrac{\partial z_1}{\partial x_2} + \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_2}\dfrac{\partial z_2}{\partial x_2} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_2}\dfrac{\partial z_2}{\partial x_2}$

We can reuse previous computations

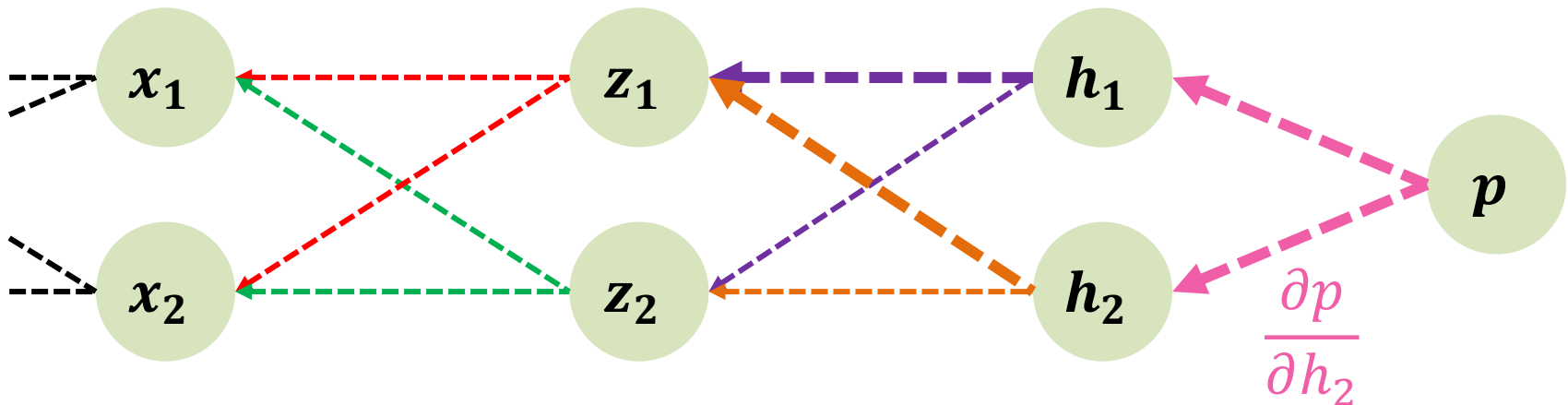3: $\dfrac{\partial p}{\partial h_1}$ $\dfrac{\partial p}{\partial h_2}$ We will need these for GD

2: $\boxed{\dfrac{\partial p}{\partial z_1} = \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_1} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_1}}$ $\dfrac{\partial p}{\partial z_2} = \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_2} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_2}$

$\dfrac{\partial p}{\partial x_1} = \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_1}\dfrac{\partial z_1}{\partial x_1} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_1}\dfrac{\partial z_1}{\partial x_1} + \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_2}\dfrac{\partial z_2}{\partial x_1} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_2}\dfrac{\partial z_2}{\partial x_1}$

1:

$\dfrac{\partial p}{\partial x_2} = \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_1}\dfrac{\partial z_1}{\partial x_2} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_1}\dfrac{\partial z_1}{\partial x_2} + \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_2}\dfrac{\partial z_2}{\partial x_2} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_2}\dfrac{\partial z_2}{\partial x_2}$

$\dfrac{\partial p}{\partial h_2}$

We can reuse previous computations

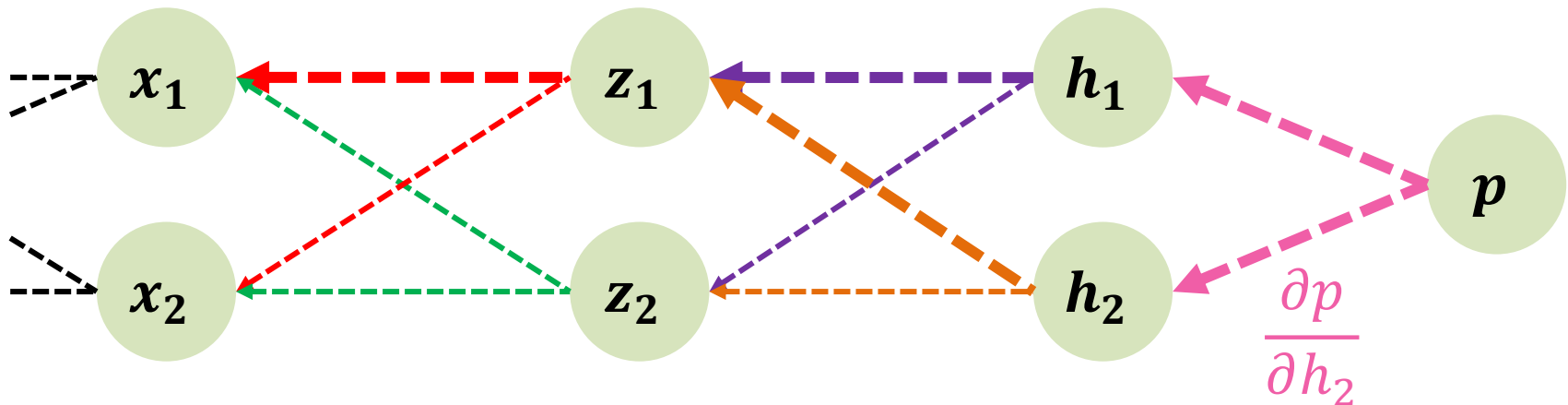3: $\quad \dfrac{\partial p}{\partial h_1} \qquad\qquad \dfrac{\partial p}{\partial h_2}$ We will need these for GD

2: $\quad \dfrac{\partial p}{\partial z_1} = \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_1} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_1} \qquad\qquad \dfrac{\partial p}{\partial z_2} = \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_2} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_2}$

$$\dfrac{\partial p}{\partial x_1} = \boxed{\dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_1}\dfrac{\partial z_1}{\partial x_1} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_1}\dfrac{\partial z_1}{\partial x_1}} + \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_2}\dfrac{\partial z_2}{\partial x_1} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_2}\dfrac{\partial z_2}{\partial x_1}$$

1:

$$\dfrac{\partial p}{\partial x_2} = \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_1}\dfrac{\partial z_1}{\partial x_2} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_1}\dfrac{\partial z_1}{\partial x_2} + \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_2}\dfrac{\partial z_2}{\partial x_2} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_2}\dfrac{\partial z_2}{\partial x_2}$$

We can reuse previous computations

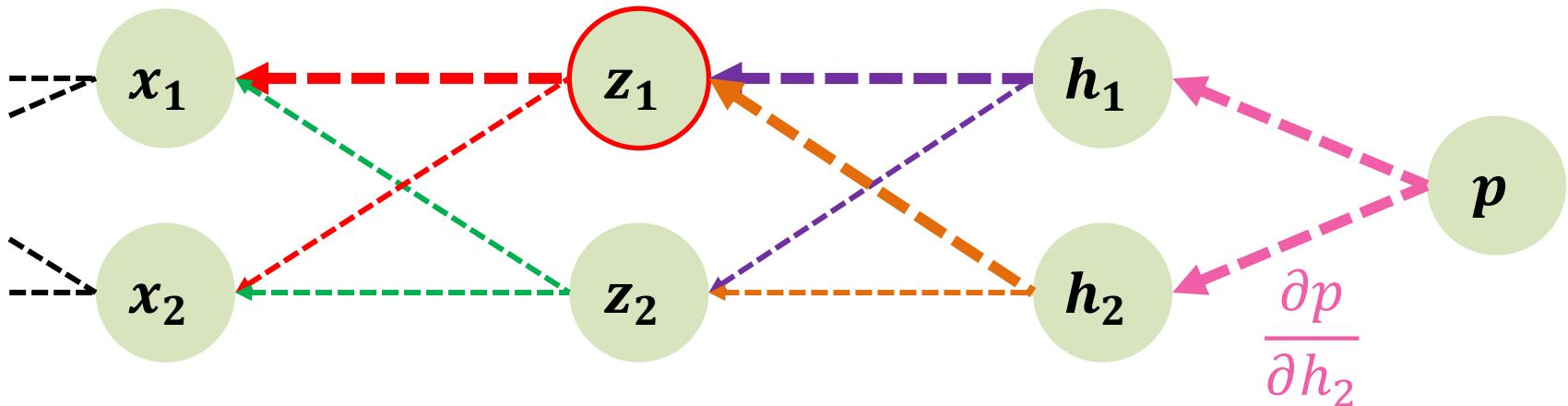3: $\quad \dfrac{\partial p}{\partial h_1} \qquad \dfrac{\partial p}{\partial h_2}$ $\qquad\qquad$ We will need these for GD

2: $\quad \dfrac{\partial p}{\partial z_1} = \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_1} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_1}$ $\qquad\qquad$ $\dfrac{\partial p}{\partial z_2} = \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_2} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_2}$

$\dfrac{\partial p}{\partial x_1} = \boxed{\left(\dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_1} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_1}\right)\dfrac{\partial z_1}{\partial x_1}} + \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_2}\dfrac{\partial z_2}{\partial x_1} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_2}\dfrac{\partial z_2}{\partial x_1}$

1:

$\dfrac{\partial p}{\partial x_2} = \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_1}\dfrac{\partial z_1}{\partial x_2} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_1}\dfrac{\partial z_1}{\partial x_2} + \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_2}\dfrac{\partial z_2}{\partial x_2} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_2}\dfrac{\partial z_2}{\partial x_2}$

$\dfrac{\partial p}{\partial h_2}$

We can reuse previous computations

We will need these for GD

3: $\dfrac{\partial p}{\partial h_1}$ $\dfrac{\partial p}{\partial h_2}$

2: $\dfrac{\partial p}{\partial z_1} = \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_1} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_1}$ $\qquad \dfrac{\partial p}{\partial z_2} = \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_2} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_2}$

$\dfrac{\partial p}{\partial x_1} = \boxed{\left(\dfrac{\partial p}{\partial z_1}\right)\dfrac{\partial z_1}{\partial x_1}} + \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_2}\dfrac{\partial z_2}{\partial x_1} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_2}\dfrac{\partial z_2}{\partial x_1}$

1:

$\dfrac{\partial p}{\partial x_2} = \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_1}\dfrac{\partial z_1}{\partial x_2} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_1}\dfrac{\partial z_1}{\partial x_2} + \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_2}\dfrac{\partial z_2}{\partial x_2} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_2}\dfrac{\partial z_2}{\partial x_2}$

$\dfrac{\partial p}{\partial h_2}$

We can reuse previous computations

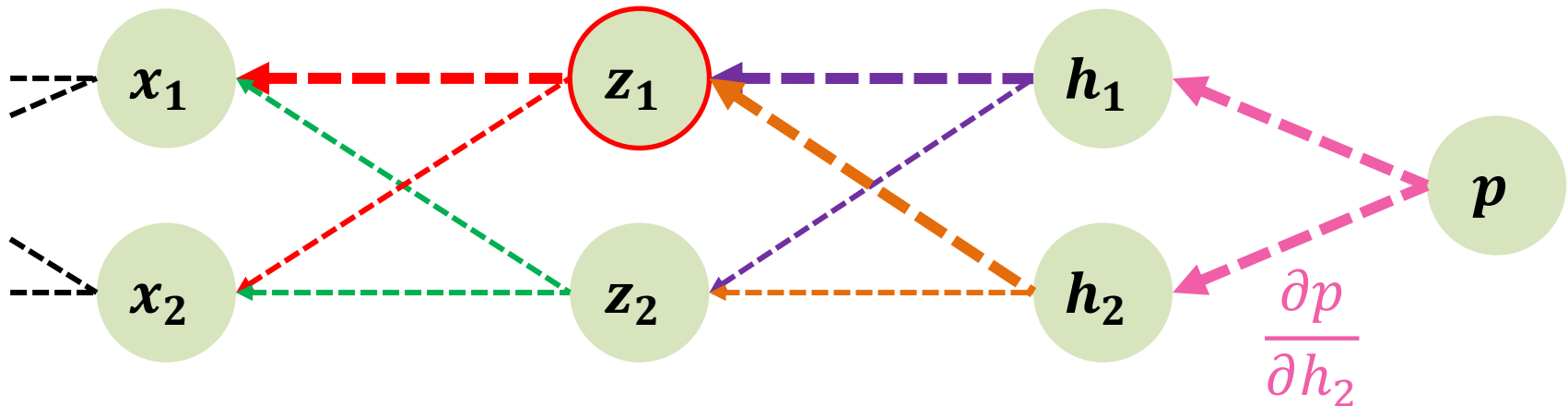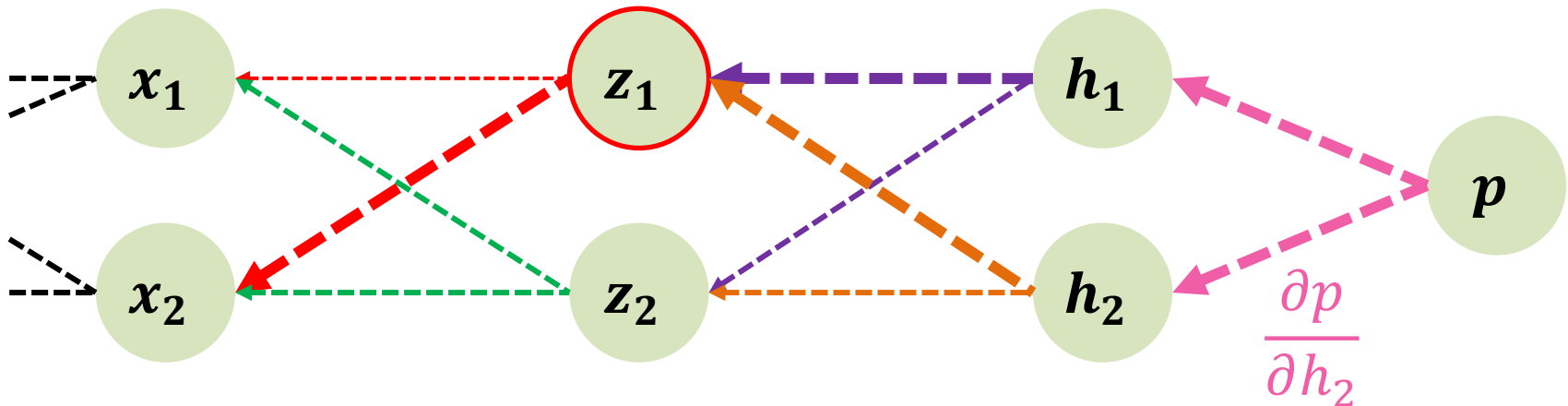3: $\quad \dfrac{\partial p}{\partial h_1} \qquad \dfrac{\partial p}{\partial h_2}$ $\qquad\qquad$ We will need these for GD

2: $\quad \dfrac{\partial p}{\partial z_1} = \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_1} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_1}$ $\qquad\qquad$ $\dfrac{\partial p}{\partial z_2} = \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_2} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_2}$

$$\dfrac{\partial p}{\partial x_1} = \left(\dfrac{\partial p}{\partial z_1}\right)\dfrac{\partial z_1}{\partial x_1} + \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_2}\dfrac{\partial z_2}{\partial x_1} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_2}\dfrac{\partial z_2}{\partial x_1}$$

1:

$$\dfrac{\partial p}{\partial x_2} = \boxed{\left(\dfrac{\partial p}{\partial z_1}\right)\dfrac{\partial z_1}{\partial x_2}} + \dfrac{\partial p}{\partial h_1}\dfrac{\partial h_1}{\partial z_2}\dfrac{\partial z_2}{\partial x_2} + \dfrac{\partial p}{\partial h_2}\dfrac{\partial h_2}{\partial z_2}\dfrac{\partial z_2}{\partial x_2}$$
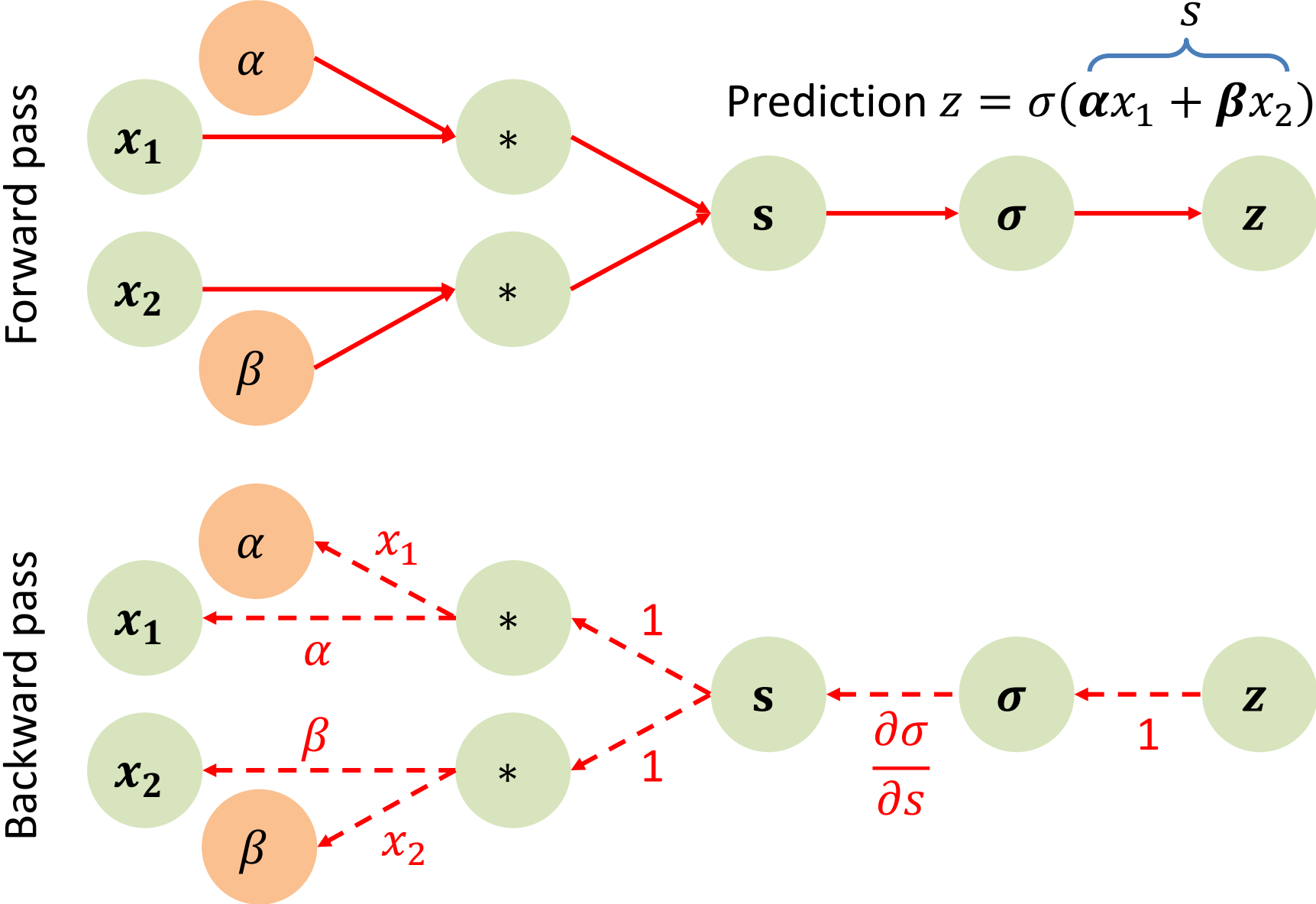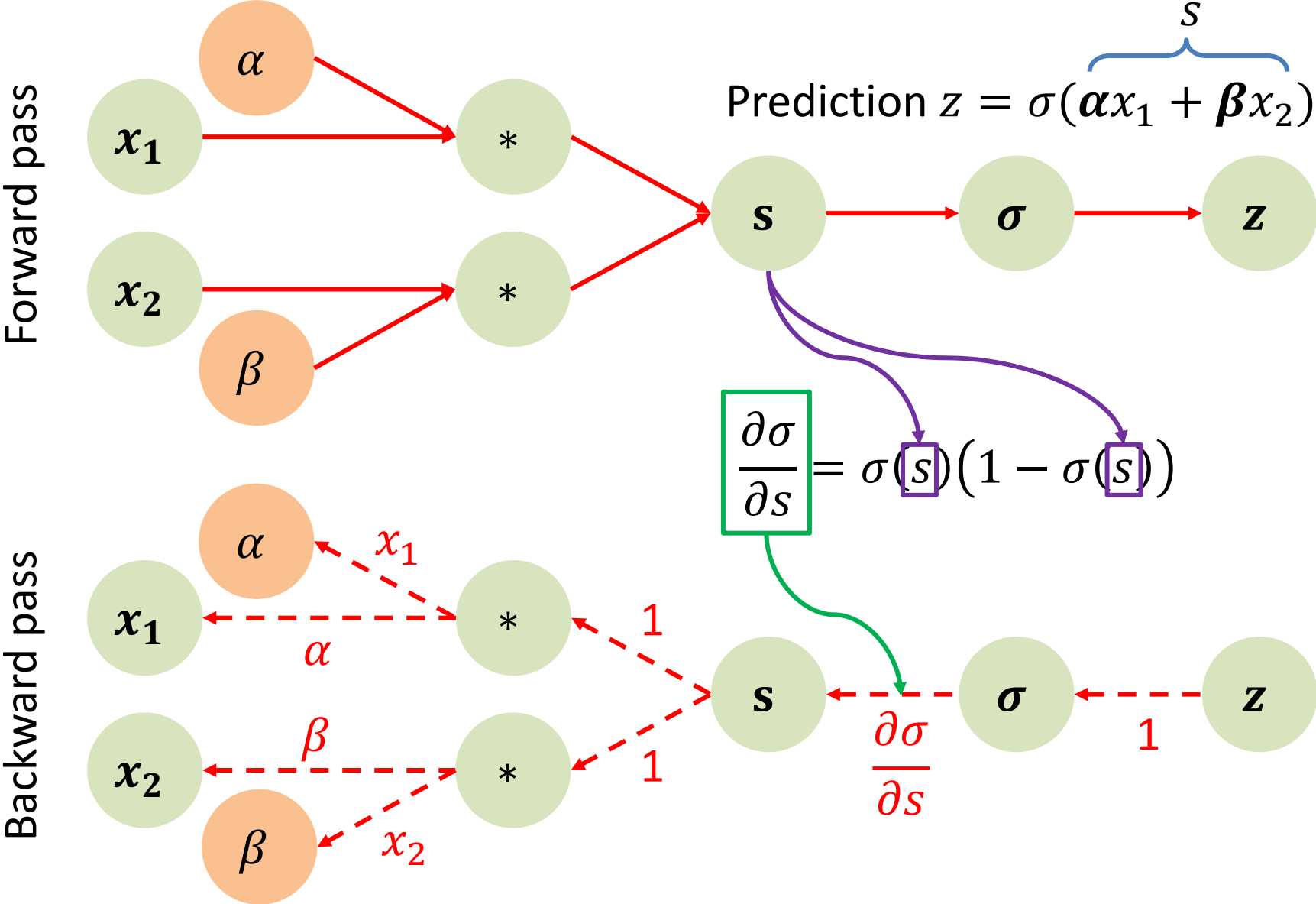
This is called reverse-mode differentiation

- In application to neural networks it has one more name: **back-propagation**.

- It works **fast**, because we reuse computations from previous steps.

- In fact, for each edge we compute its value only once. And multiply by its value exactly once.

Back-propagation (Back-prop)

Forward pass

Prediction $z = \sigma(\boldsymbol{\alpha} x_1 + \boldsymbol{\beta} x_2)$

Backward pass

Back-propagation (Back-prop)

Prediction $z = \sigma(\boldsymbol{\alpha} x_1 + \boldsymbol{\beta} x_2)$

$$\frac{\partial \sigma}{\partial s} = \sigma(s)\big(1 - \sigma(s)\big)$$

Forward pass interface

Let's implement a sigmoid activation node!

Prediction $z = \sigma(\boldsymbol{\alpha} x_1 + \boldsymbol{\beta} x_2)$
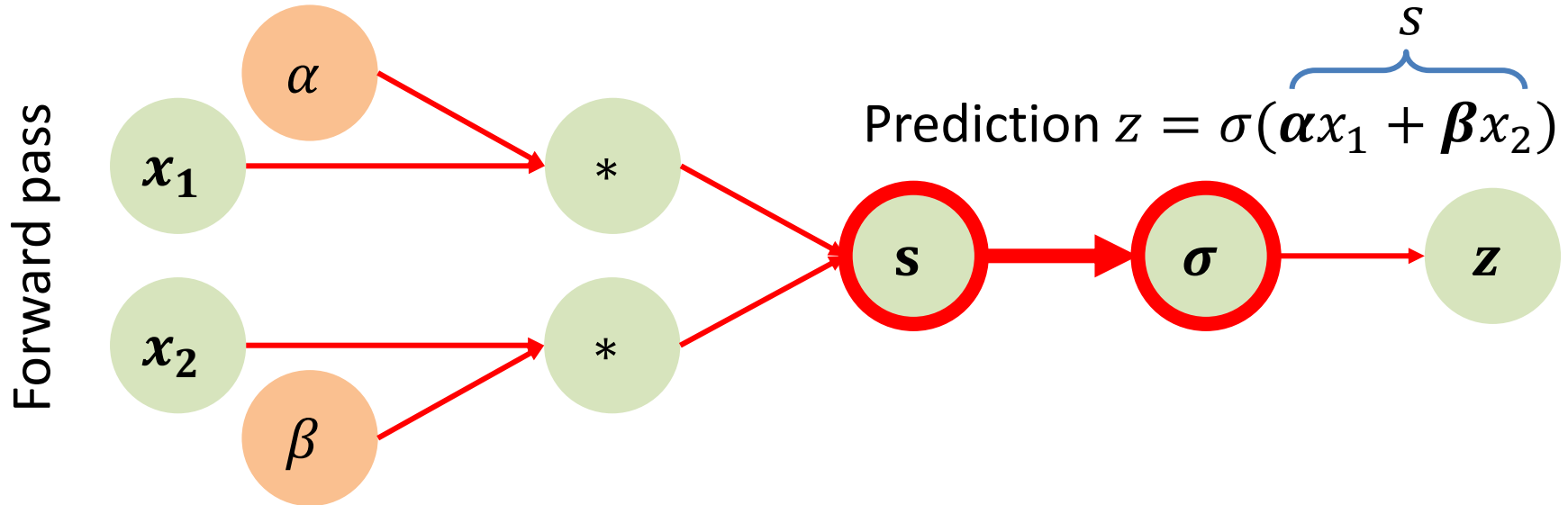
```python
def forward_pass(inputs):
    return 1. / (1 + np.exp(-inputs))
```
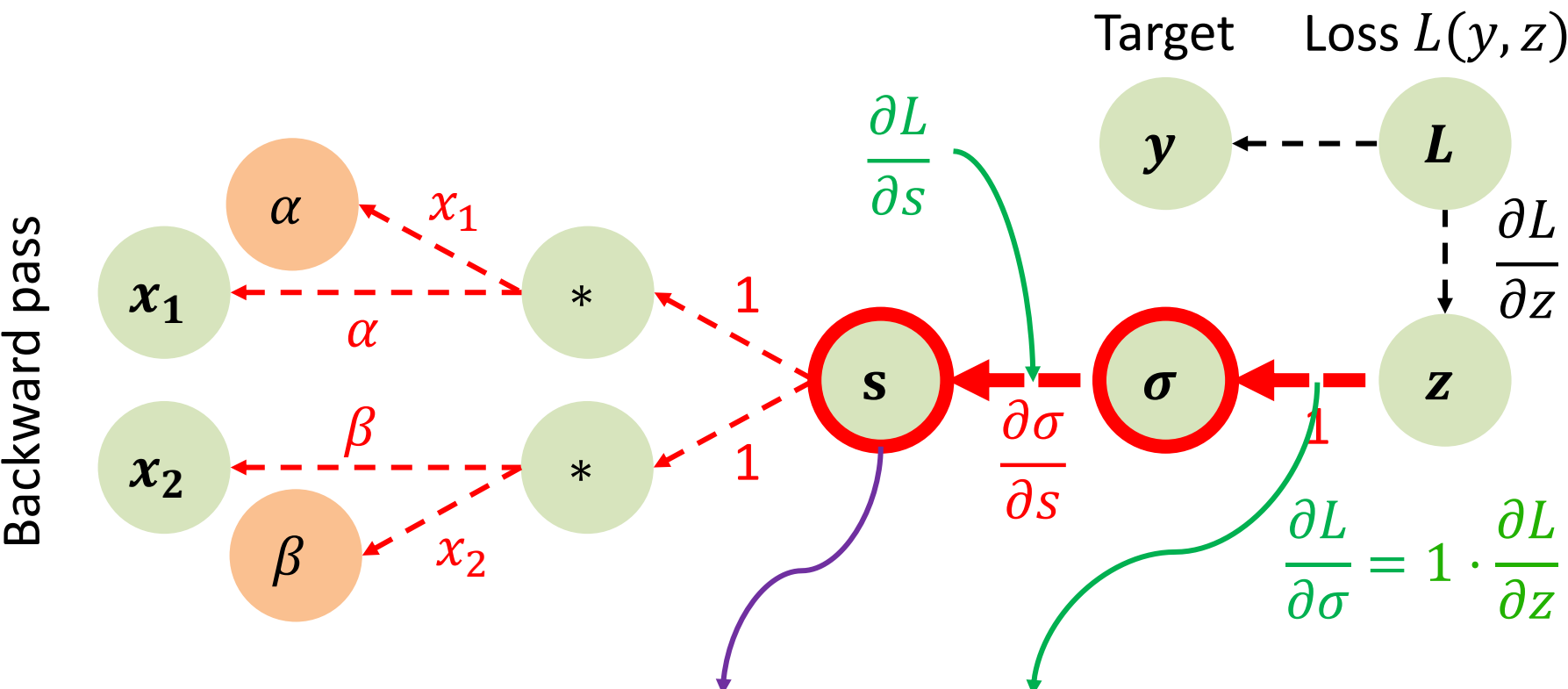
Backward pass interface

Backward pass interface

Target   Loss $L(y, z)$

y ⟵ L

$\dfrac{\partial L}{\partial s}$

$\dfrac{\partial L}{\partial z}$

Backward pass

α

x_1 ⟷ x_1 / α · *

x_2 · β · *

β · x_2

s ⟵ σ ⟵ z

$\dfrac{\partial \sigma}{\partial s}$ 1

$\dfrac{\partial L}{\partial \sigma} = 1 \cdot \dfrac{\partial L}{\partial z}$
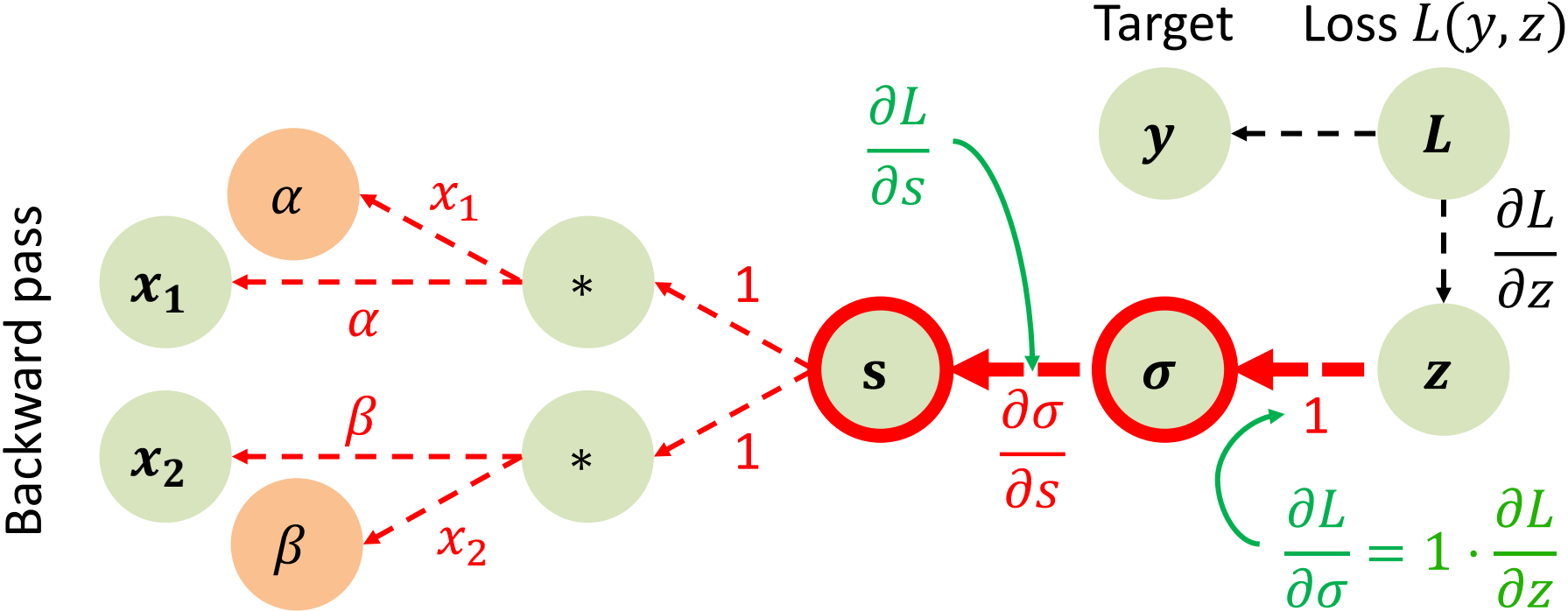
```
def backward_pass(inputs, incoming_gradient):
    sigmoid = 1. / (1 + np.exp(-inputs))
    return sigmoid * (1 - sigmoid) * incoming_gradient
```

$$\underbrace{\dfrac{\partial L}{\partial s}}_{} \quad = \quad \underbrace{\dfrac{\partial \sigma}{\partial s}}_{} \quad \cdot \quad \dfrac{\partial L}{\partial \sigma}$$

Summary

- Back-propagation is a name for an **automatic** reverse-mode differentiation

- Back-propagation is **fast**

- And this is how neural nets are trained today

- In the next video we will look at MLP implementation details