Intro

- In this video we will learn how to compute the gradients for MLP **automatically**.

Chain rule

- We know derivatives for simple functions:

$$\frac{dx^2}{dx} = 2x \qquad \frac{de^x}{dx} = e^x \qquad \frac{d\ln(x)}{dx} = \frac{1}{x}$$

- Let's take a composite function:

$$z_1 = z_1(x_1, x_2)$$

$$z_2 = z_2(x_1, x_2)$$

where z_1, z_2, p are differentiable

$$p = p(z_1, z_2)$$

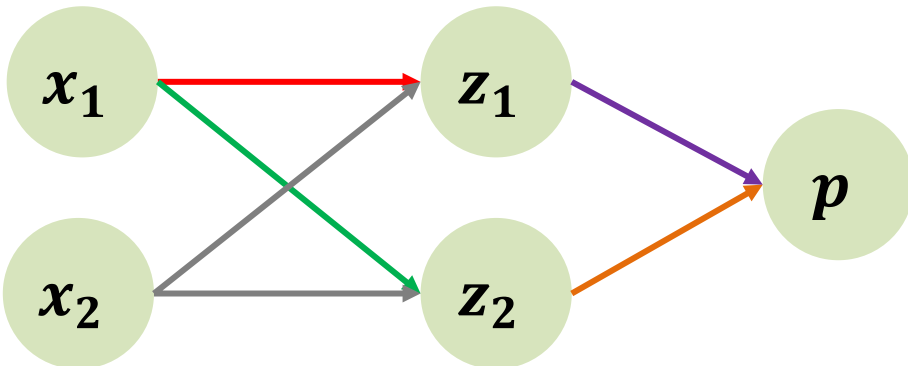
Chain rule: $$\frac{\partial p}{\partial x_1} = \frac{\partial p}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial p}{\partial z_2}\frac{\partial z_2}{\partial x_1}$$

Example for $h(x) = f(x)g(x)$:

$$\frac{\partial h}{\partial x} = \frac{\partial h}{\partial f}\frac{\partial f}{\partial x} + \frac{\partial h}{\partial g}\frac{\partial g}{\partial x} = g\frac{\partial f}{\partial x} + f\frac{\partial g}{\partial x}$$

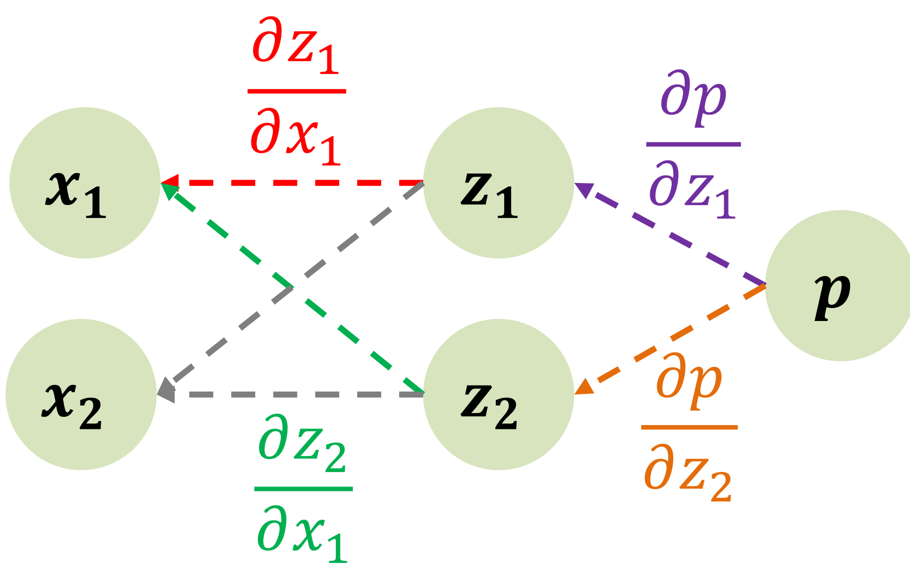
Derivatives computation graph

- Let's take our simple computation graph:



$$z_1 = z_1(x_1, x_2)$$
$$z_2 = z_2(x_1, x_2)$$
$$p = p(z_1, z_2)$$

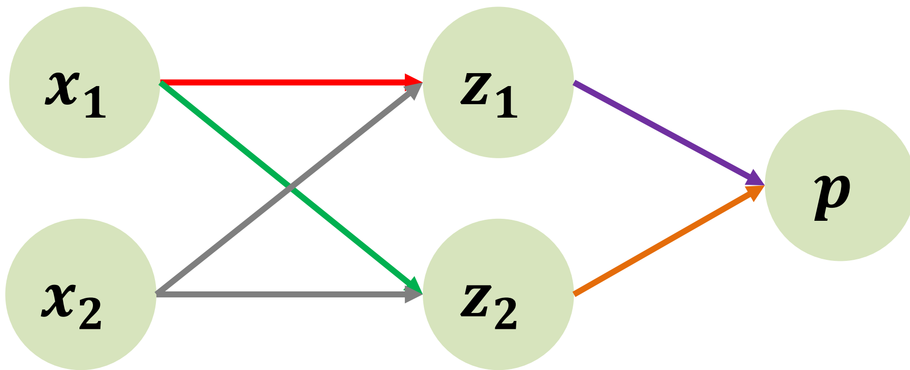
- And construct a new graph of derivatives:



Each edge is assigned to derivative of origin w.r.t. destination

Derivatives computation graph

- Let’s take our simple computation graph:

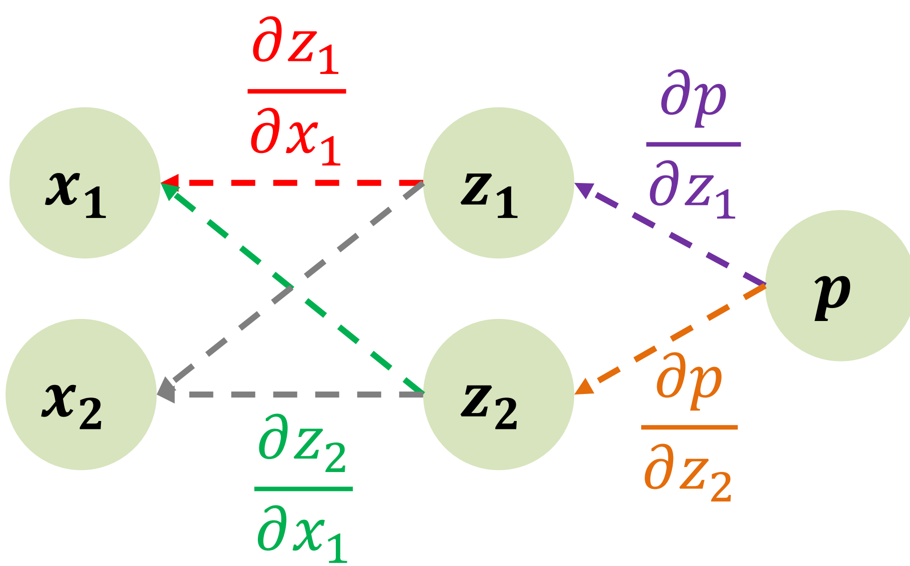


$$z_1 = z_1(x_1, x_2)$$

$$z_2 = z_2(x_1, x_2)$$

$$p = p(z_1, z_2)$$

- And construct a new graph of derivatives:

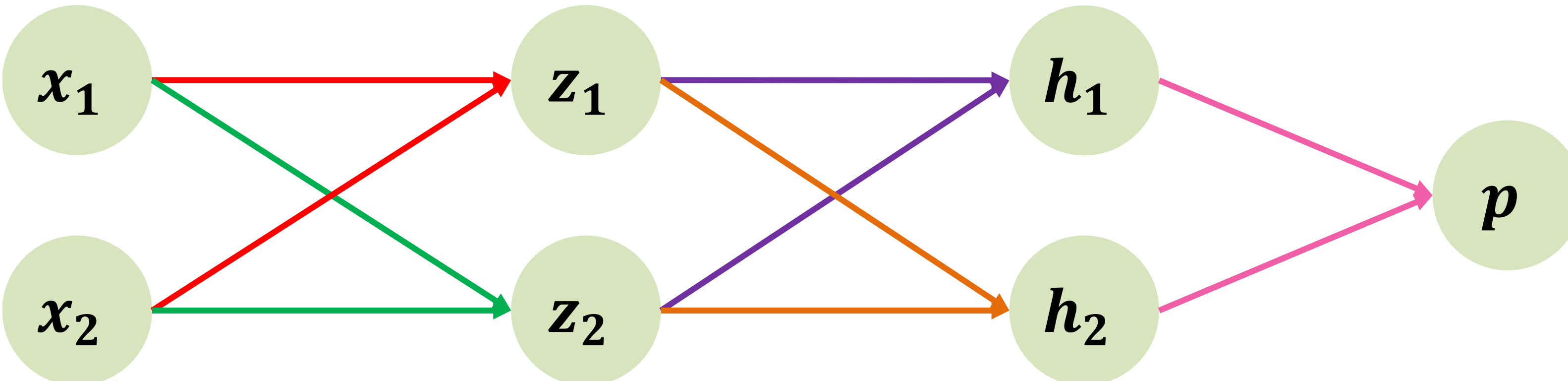


$$\frac{\partial p}{\partial x_1} = \frac{\partial p}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial p}{\partial z_2}\frac{\partial z_2}{\partial x_1}$$

You can see how
a **chain rule** works

Let's go deeper

- A little bit more composite function:



$$z_1 = z_1(x_1, x_2) \qquad h_1 = h_1(z_1, z_2)$$

$$z_2 = z_2(x_1, x_2) \qquad h_2 = h_2(z_1, z_2)$$

$$p = p(h_1, h_2)$$

Let's go deeper

Chain rule: $$\frac{\partial p}{\partial x_1} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial x_1} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial x_1}$$

Let's go deeper

Chain rule:
$$\frac{\partial p}{\partial x_1} = \frac{\partial p}{\partial h_1}\boxed{\frac{\partial h_1}{\partial x_1}} + \frac{\partial p}{\partial h_2}\boxed{\frac{\partial h_2}{\partial x_1}}$$

$$\boxed{\frac{\partial h_1}{\partial x_1} = \frac{\partial h_1}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial h_1}{\partial z_2}\frac{\partial z_2}{\partial x_1}}$$

$$\boxed{\frac{\partial h_2}{\partial x_1} = \frac{\partial h_2}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial h_2}{\partial z_2}\frac{\partial z_2}{\partial x_1}}$$

Let's go deeper

Chain rule: $$\frac{\partial p}{\partial x_1} = \frac{\partial p}{\partial h_1}\boxed{\frac{\partial h_1}{\partial x_1}} + \frac{\partial p}{\partial h_2}\boxed{\frac{\partial h_2}{\partial x_1}}$$

these are all the possible ways to get from h1 to x1. >>

$$\boxed{\frac{\partial h_1}{\partial x_1} = \frac{\partial h_1}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial h_1}{\partial z_2}\frac{\partial z_2}{\partial x_1}}$$

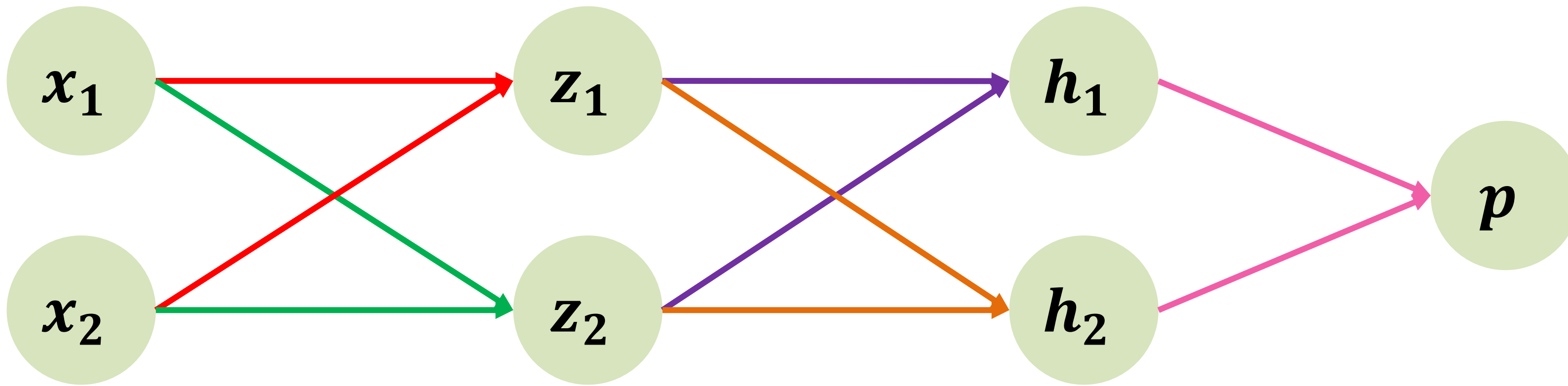
$$\boxed{\frac{\partial h_2}{\partial x_1} = \frac{\partial h_2}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial h_2}{\partial z_2}\frac{\partial z_2}{\partial x_1}}$$

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

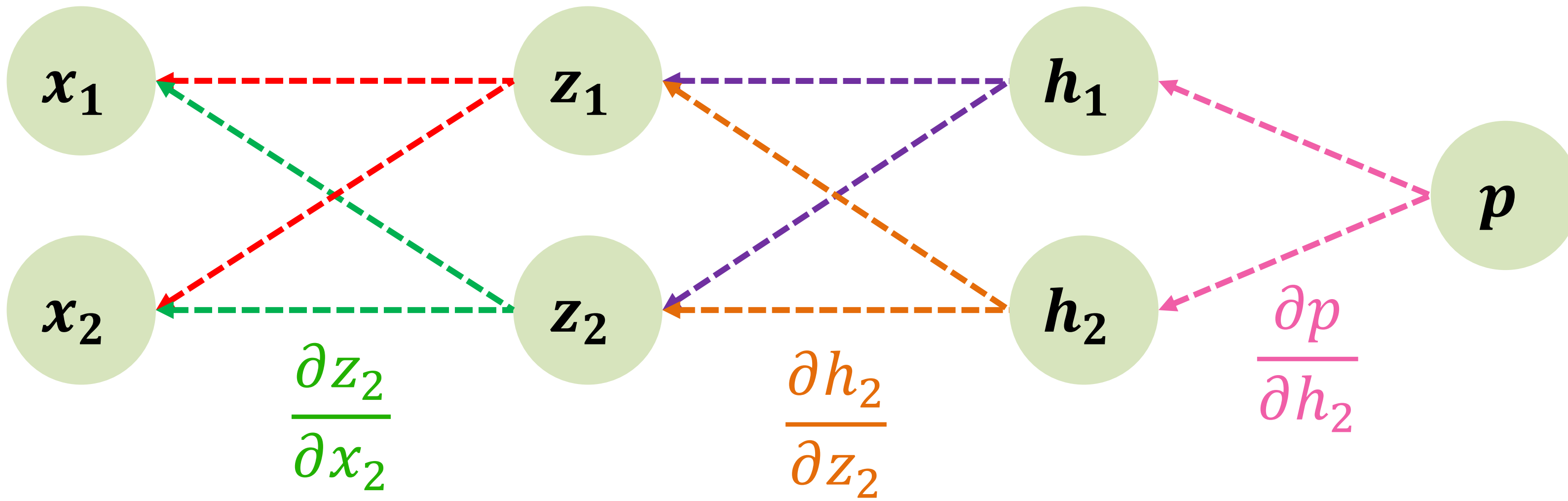
$$\frac{\partial p}{\partial x_1} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_2}\frac{\partial z_2}{\partial x_1} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_1}\frac{\partial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

Let's check out the derivatives graph!

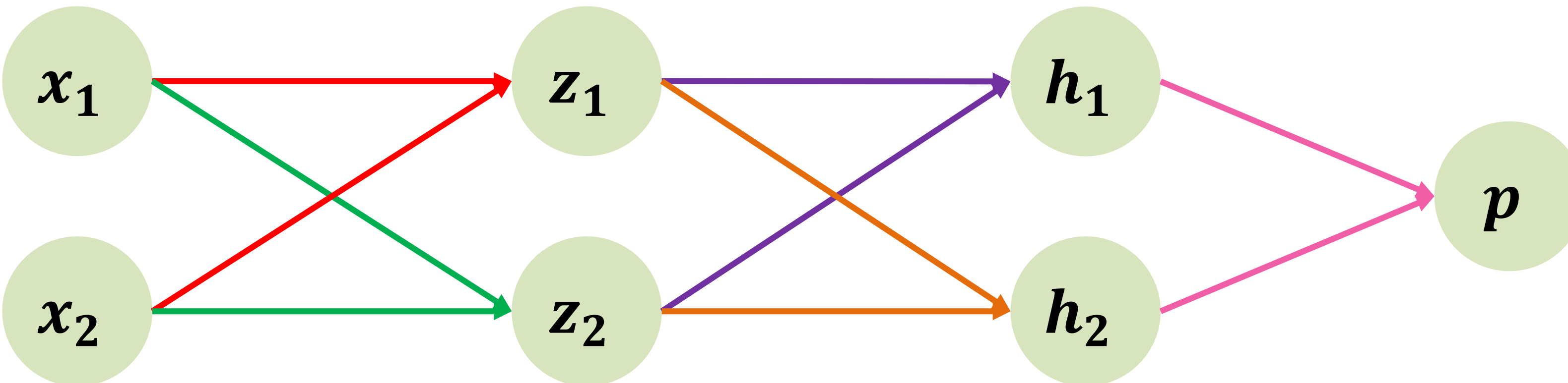
Let's go deeper



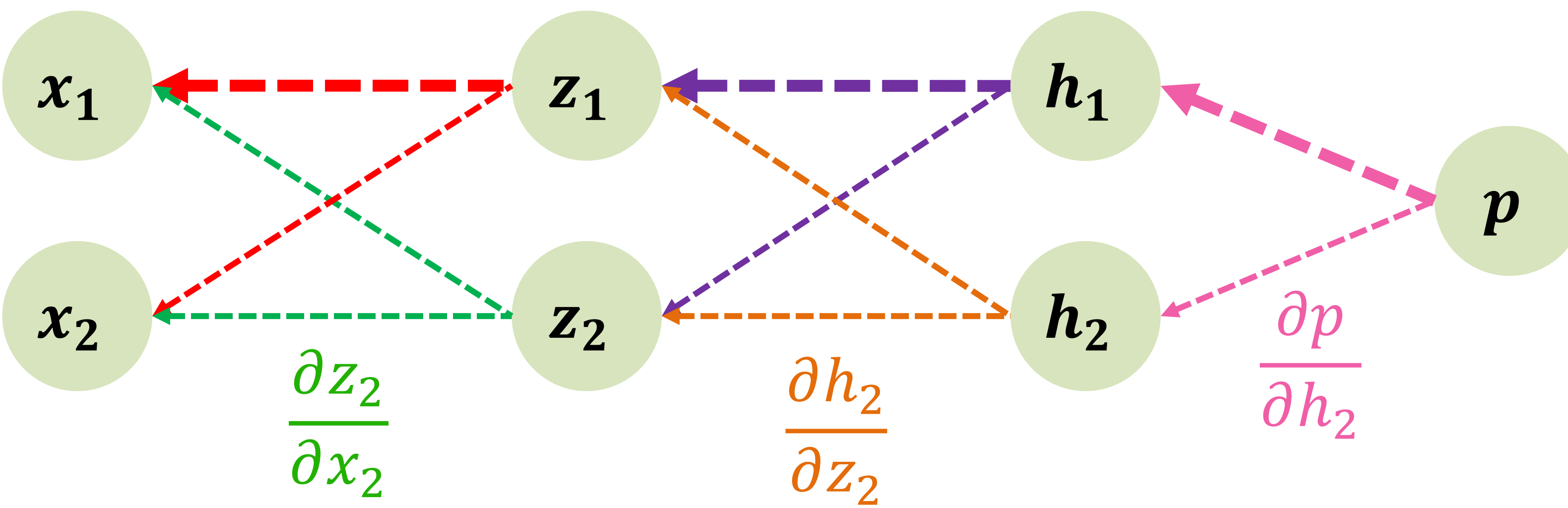
$$\frac{\partial p}{\partial x_1} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_2}\frac{\partial z_2}{\partial x_1} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_1}\frac{\partial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



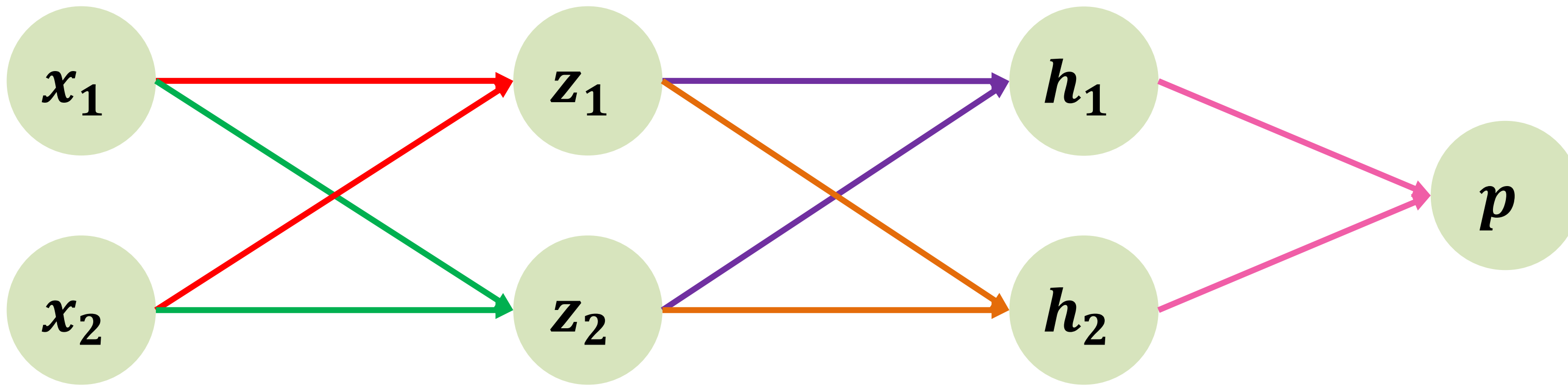
Let's go deeper



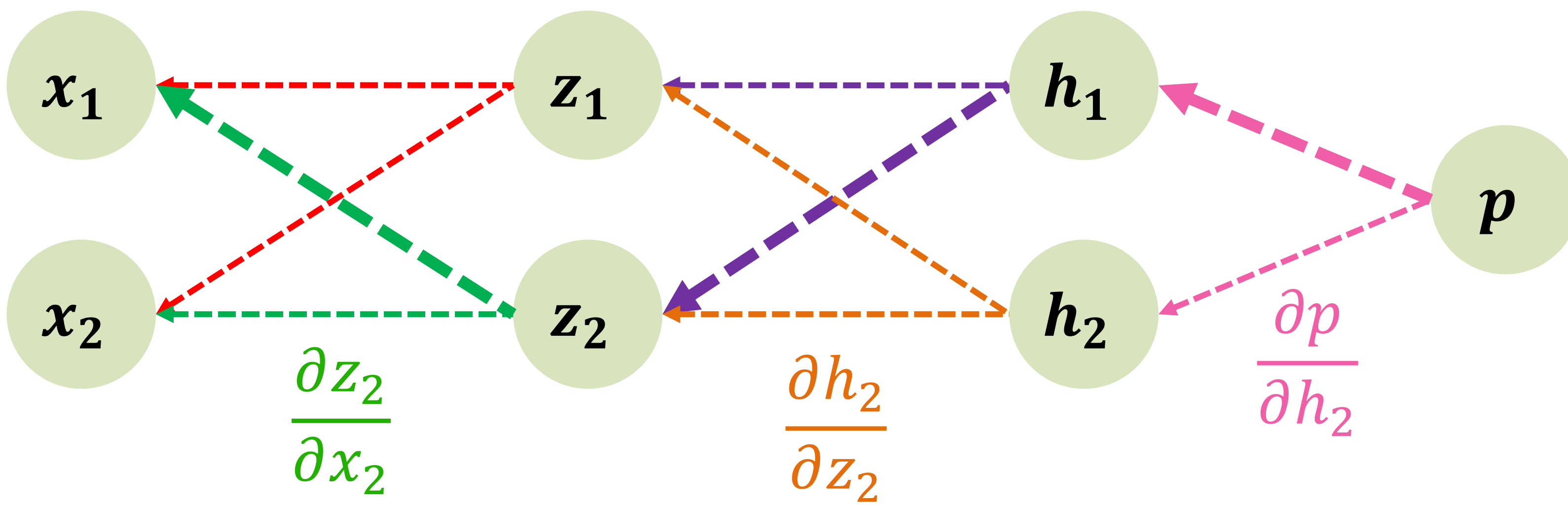
$$\frac{\partial p}{\partial x_1} = \boxed{\frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_1}\frac{\partial z_1}{\partial x_1}} + \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_2}\frac{\partial z_2}{\partial x_1} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial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



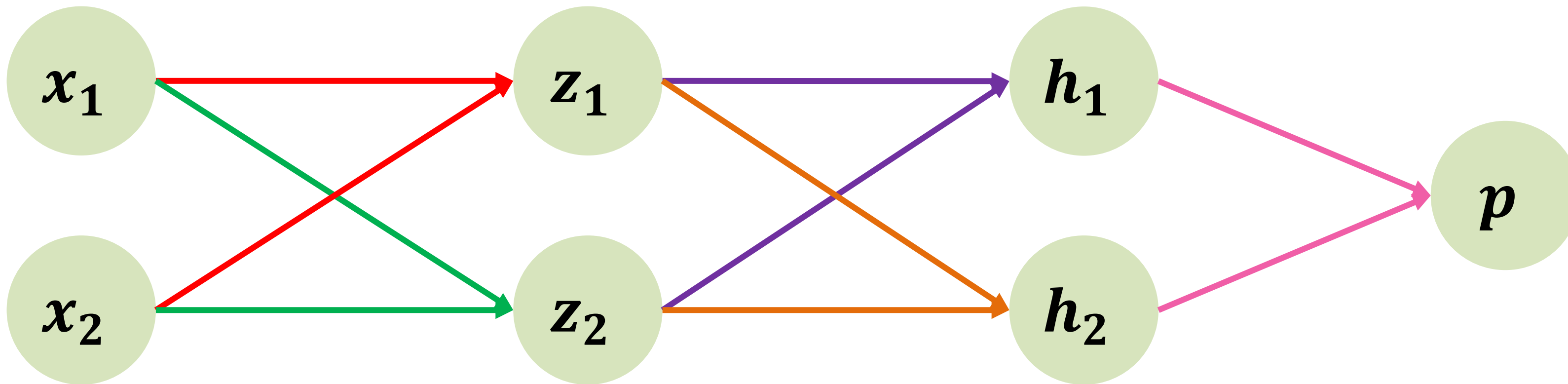
Let’s go deeper



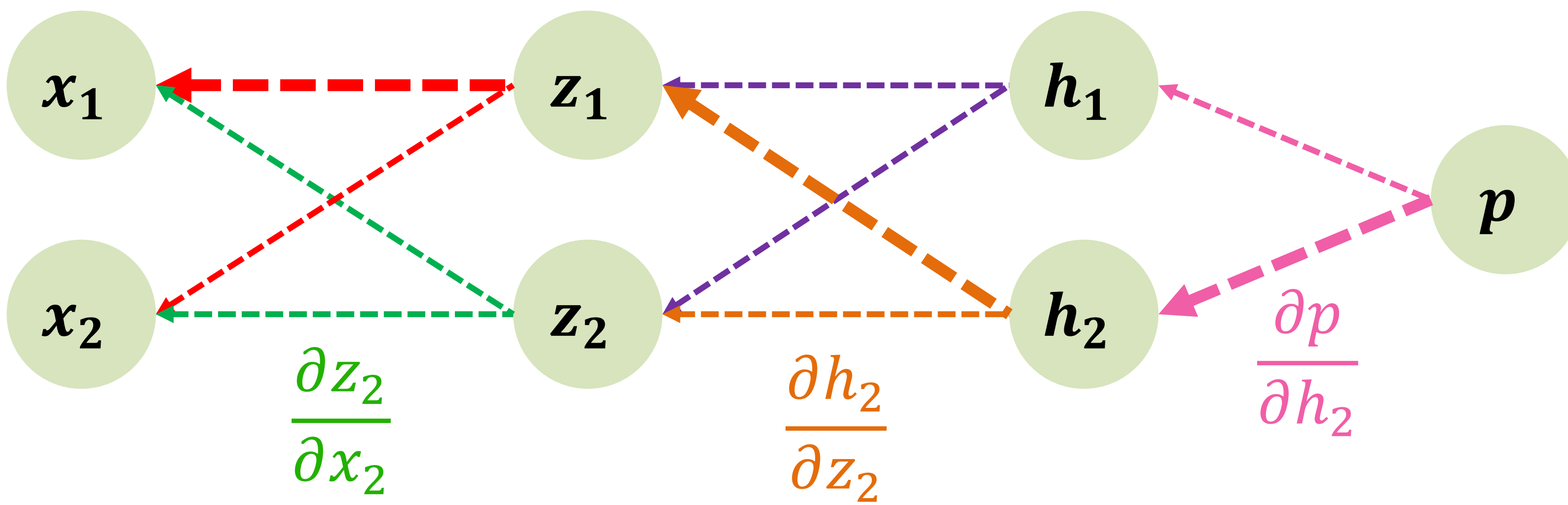
$$\frac{\partial p}{\partial x_1} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \boxed{\frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_2}\frac{\partial z_2}{\partial x_1}} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial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



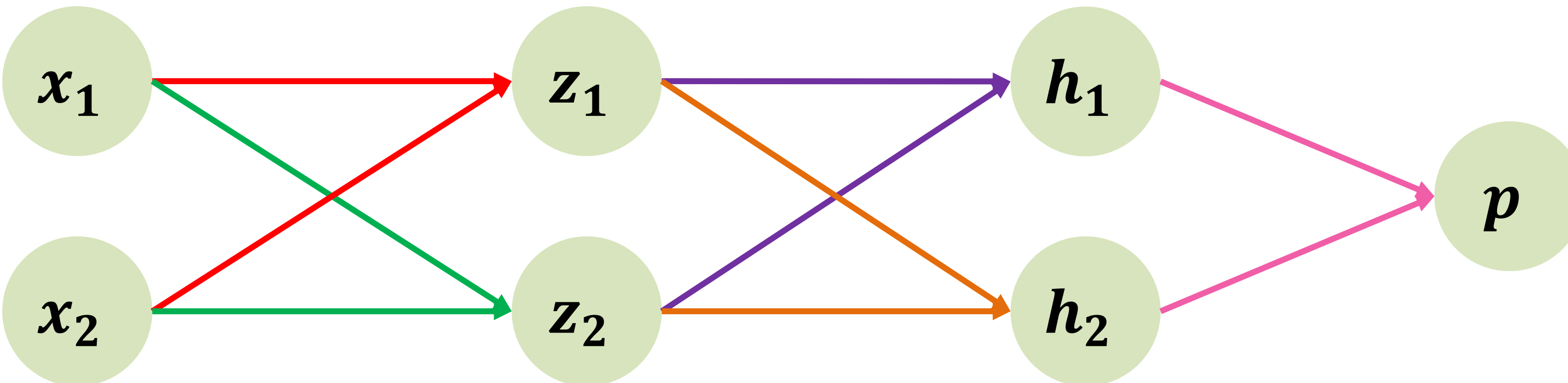
Let’s go deeper



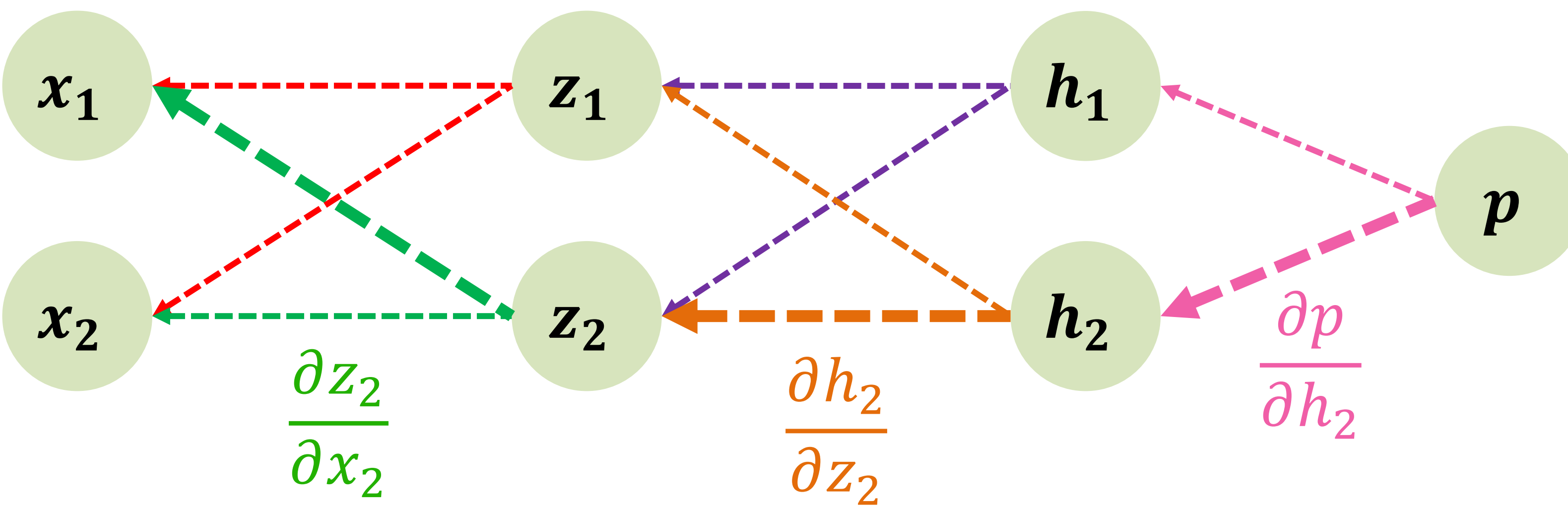
$$\frac{\partial p}{\partial x_1} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_2}\frac{\partial z_2}{\partial x_1} + \boxed{\frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial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



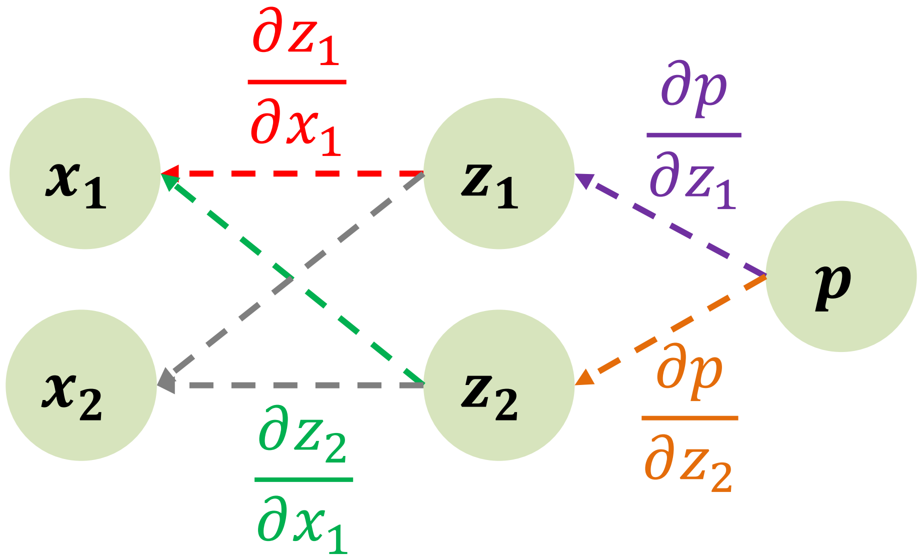
Let’s go deeper



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



How this graph of derivatives helps

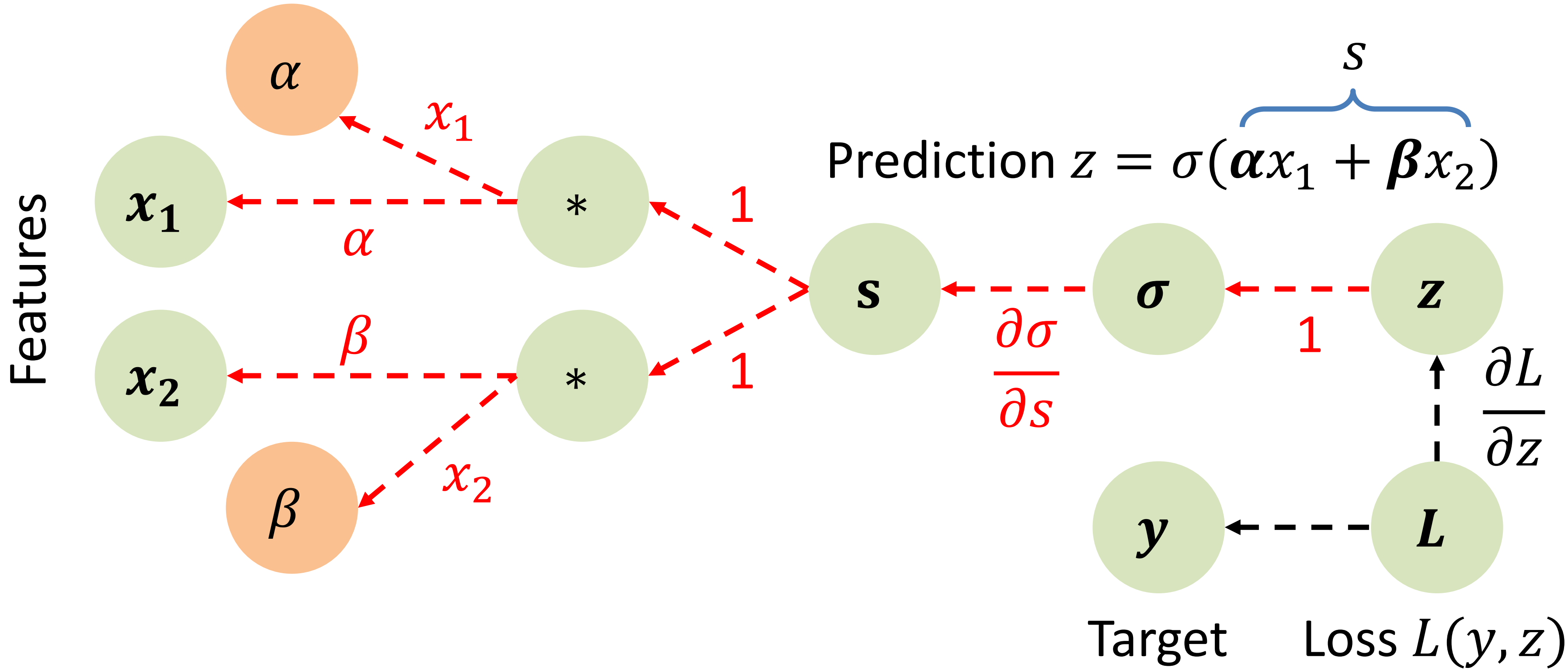


$$\frac{\partial p}{\partial x_1} = \frac{\partial p}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial p}{\partial z_2}\frac{\partial z_2}{\partial x_1}$$

How to calculate a derivative of node a w.r.t. node b:

- Find an unvisited path from a to b
- Multiply all edge values along this path
- Add to the resulting derivative

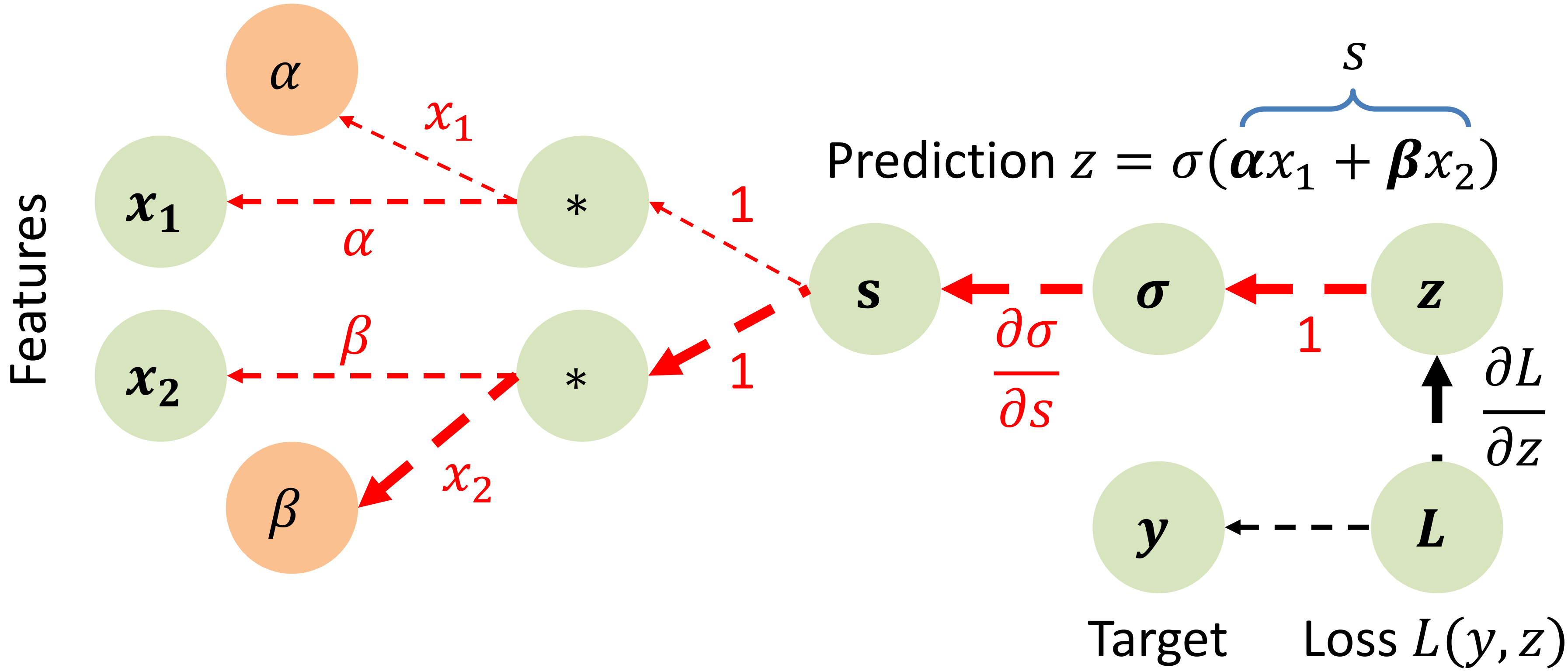
Derivatives computation graph



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

For SGD to work we need $\frac{\partial L}{\partial \alpha}, \frac{\partial L}{\partial \beta}$

Derivatives computation graph



For SGD to work we need $\dfrac{\partial L}{\partial \alpha}, \dfrac{\partial L}{\partial \beta}$

NOTE: z is basically sigma. See the '1' above.
so dsigma / ds = dz / ds

x_1 = ds / da
x_2 = ds/ dbeta

$$\frac{\partial L}{\partial \alpha} = \frac{\partial L}{\partial z} \frac{\partial \sigma}{\partial s} x_1 \qquad \frac{\partial L}{\partial \beta} = \frac{\partial L}{\partial z} \frac{\partial \sigma}{\partial s} x_2$$

Summary

- We can use chain rule to compute derivatives of composite functions
- We can use a computation graph of derivatives to compute them **automatically**
- In the next video we will find out how to do this **fast!**