

1.3 Backpropagation

作用：compute the gradient efficiently

原理：导数的链式法则

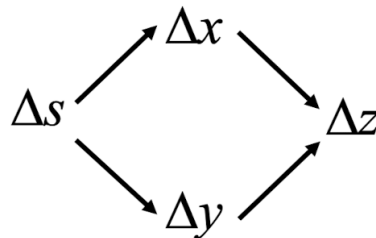
Case 1 $y = g(x) \quad z = h(y)$

$$\Delta x \rightarrow \Delta y \rightarrow \Delta z$$

$$\frac{dz}{dx} = \frac{dz}{dy} \frac{dy}{dx}$$

Case 2

$$x = g(s) \quad y = h(s) \quad z = k(x, y)$$



$$\frac{dz}{ds} = \frac{\partial z}{\partial x} \frac{dx}{ds} + \frac{\partial z}{\partial y} \frac{dy}{ds}$$

只关注某一笔 data 的偏微分

以 weight 为例, b 同理 | 对每个 x^n 都有 C^n 定义 y^n 与 \hat{y}^n 距离的函数 C^n , 越小越好

Backpropagation

之前所有 training data 的某个 loss 值的和 上标不是几次方

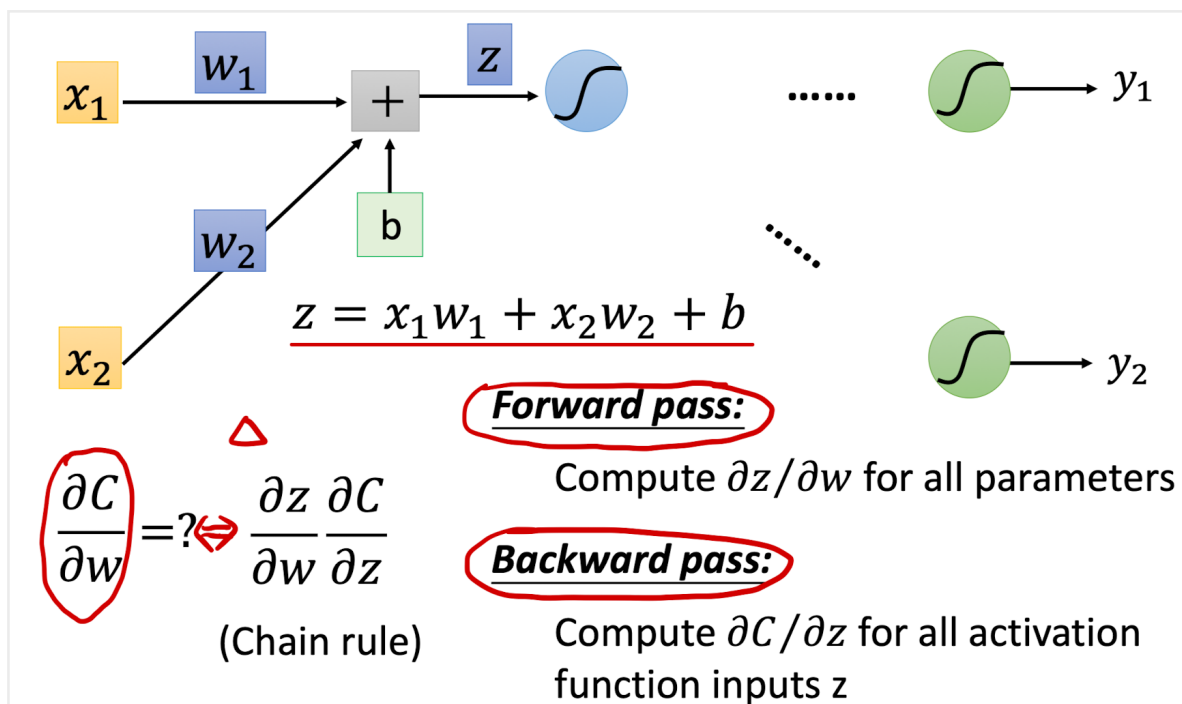
$$L(\theta) = \sum_{n=1}^N C^n(\theta)$$

对 w 偏微分 \rightarrow

$$\frac{\partial L(\theta)}{\partial w} = \sum_{n=1}^N \frac{\partial C^n(\theta)}{\partial w}$$

只 focus 对每一笔 data 的偏微分

以下图的 neural 为例, C 对 w 的偏微分可以用链式法则一拆分成两项, 前项称作 Forward pass, 后项称作 Backward pass。



Forward pass

w₁前连接的是什么,微分后就是什么

$\frac{\partial z}{\partial w_1} = ? \quad x_1$
 $\frac{\partial z}{\partial w_2} = ? \quad x_2$

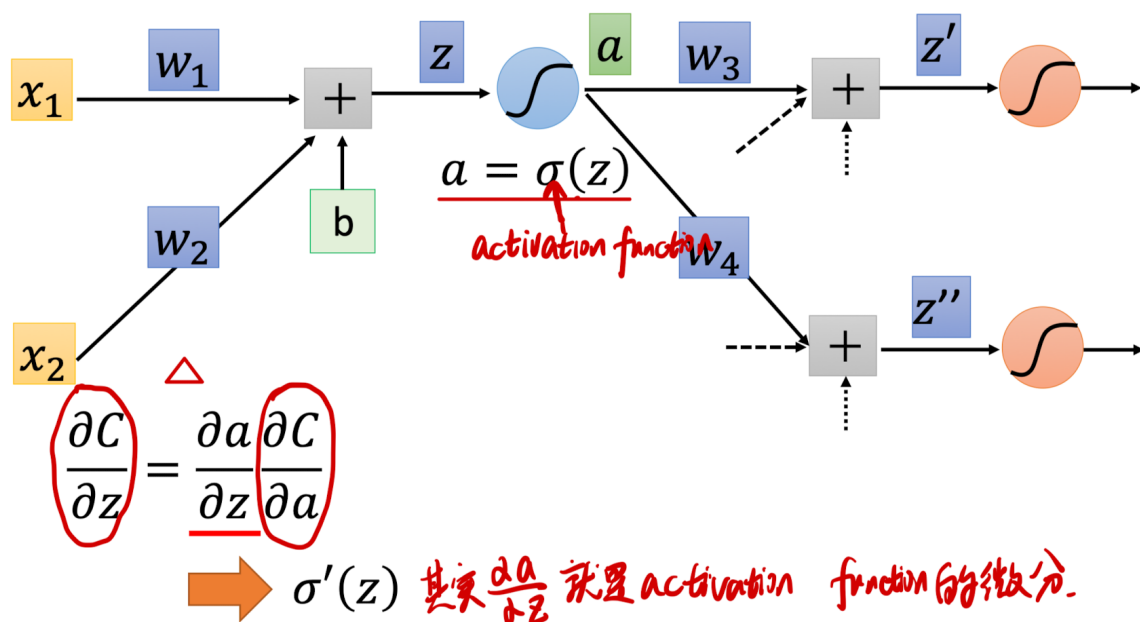
The value of the input connected by the weight

通过这样，可以算出每个参数的偏微分并记录。

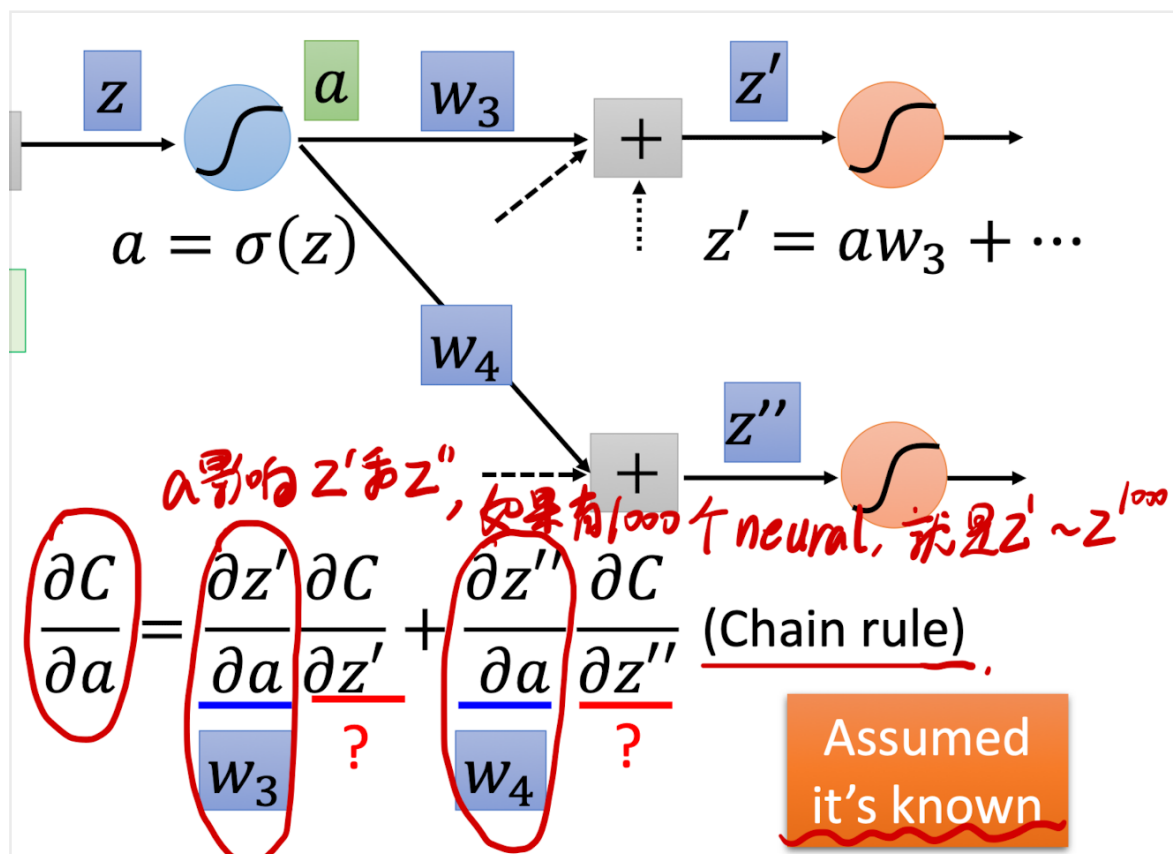
Backward pass

再用链式法则一，假设 activation function 是 sigmoid function

Compute $\partial C / \partial z$ for all activation function inputs z



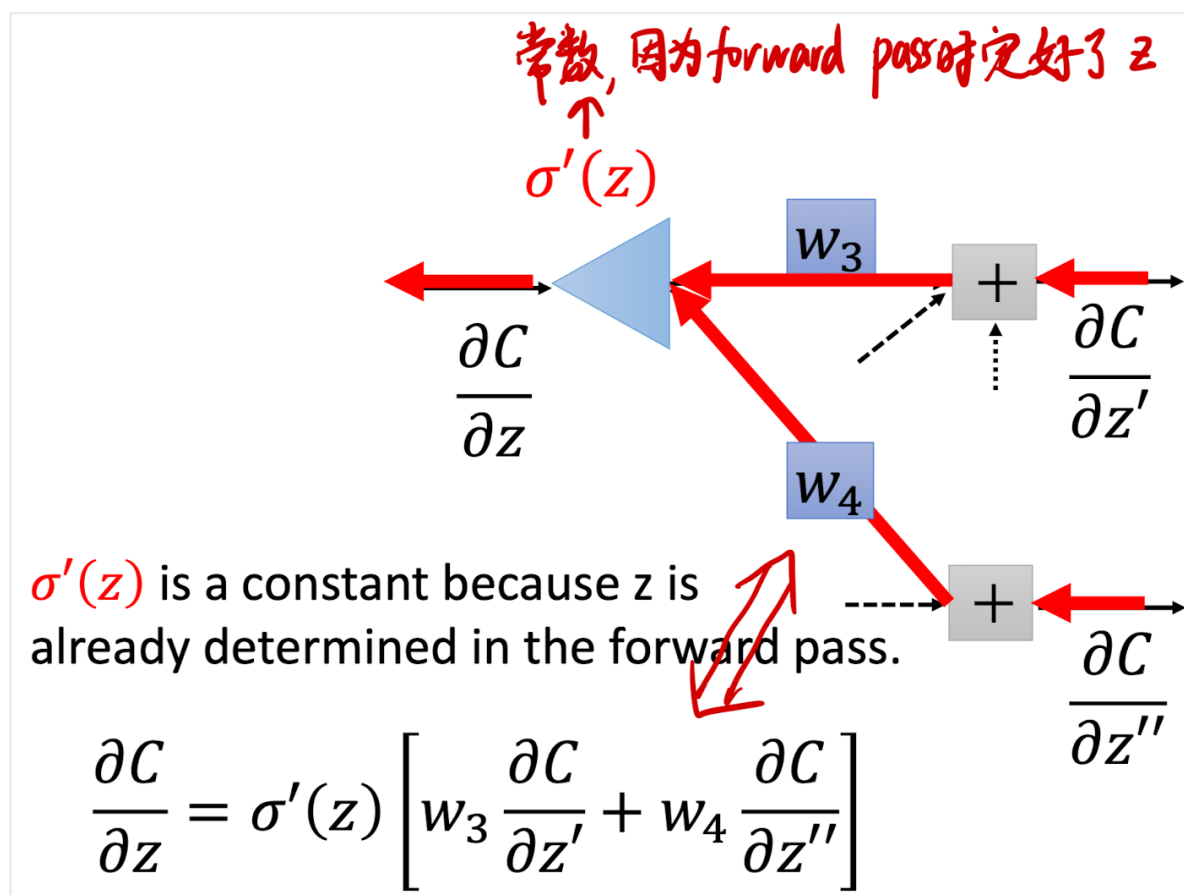
C 对 a 的偏导数，用链式法则二，并假设 C 对 z 的偏导是知道的



假设知道 $\frac{\partial C}{\partial z'}$ 和 $\frac{\partial C}{\partial z''}$, 则会算 $\frac{\partial C}{\partial z}$

$$\frac{\partial C}{\partial z} = \sigma'(z) \left[w_3 \frac{\partial C}{\partial z'} + w_4 \frac{\partial C}{\partial z''} \right]$$

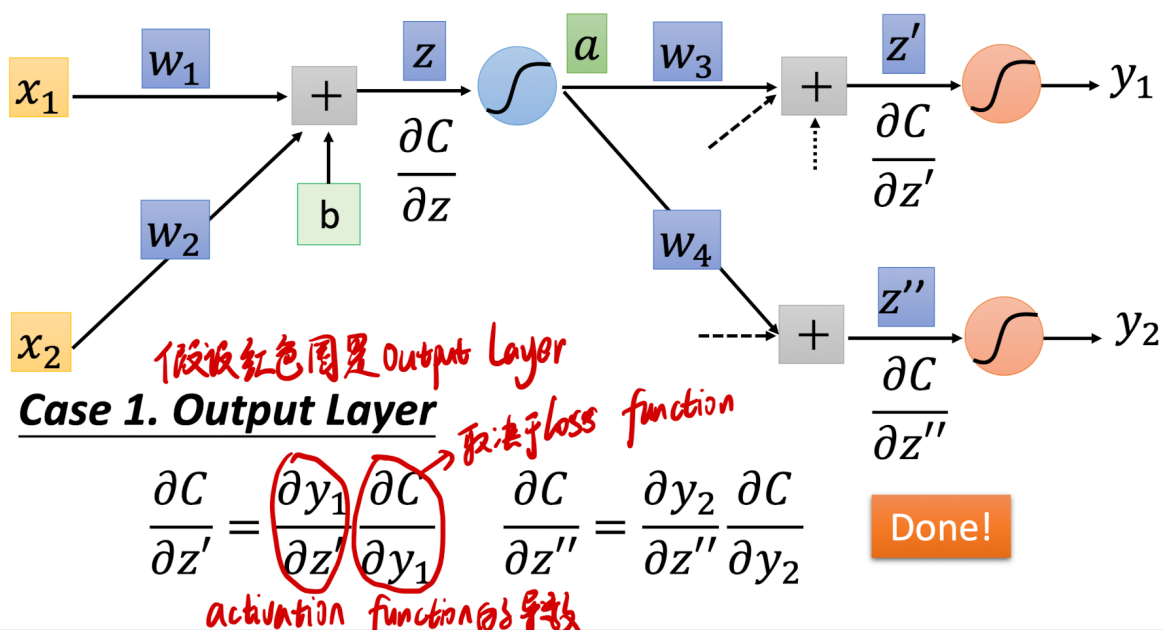
现在想象有另外一个 neural, 如下图三角形所示, input 是 C 对 z' 和 z'' 的偏导, 图中的 neural 做的事情和图中的公式是等价的。其中乘法左边一项是常数



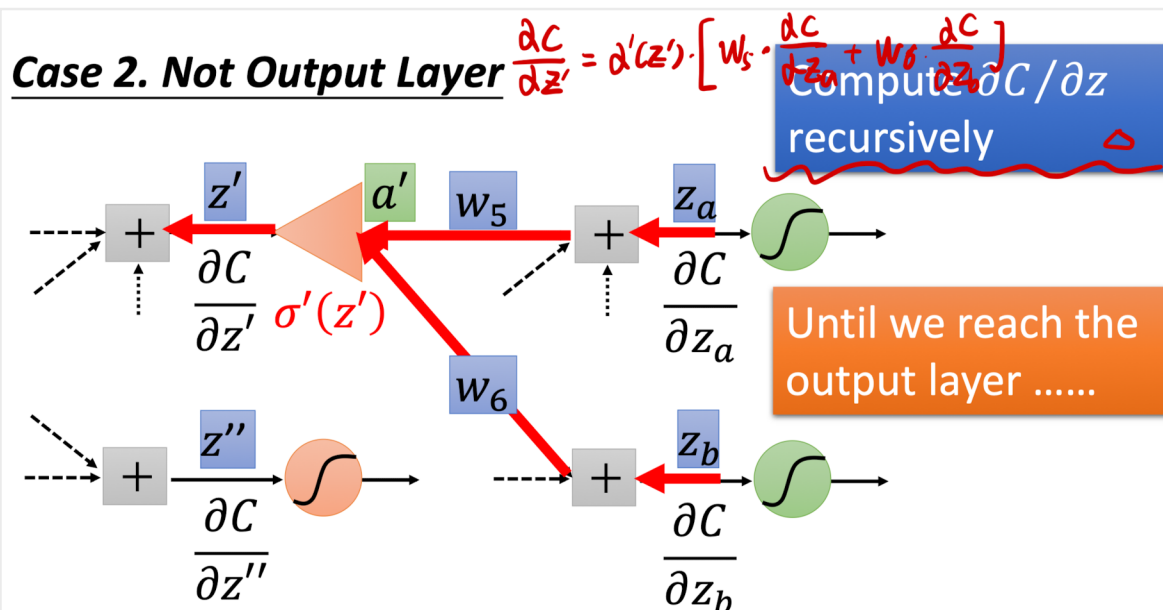
所以问题就在于如何算 C 对 z' 的偏导和 C 对 z'' 的偏导。分两种情况讨论：

Case1: 假设红色 neural 是 output layer

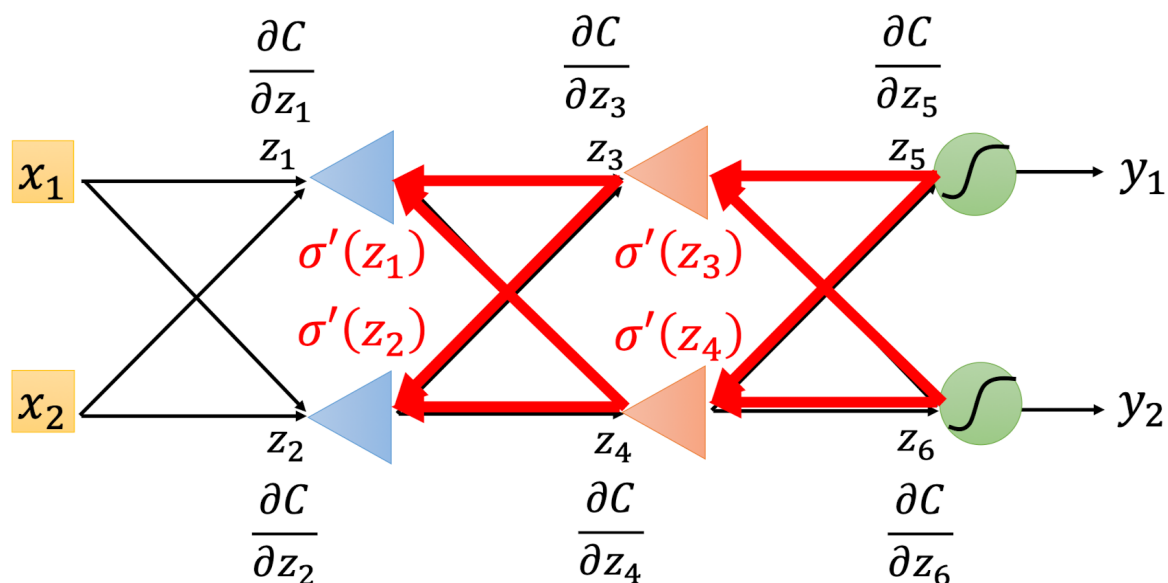
Compute $\partial C / \partial z$ for all activation function inputs z



Case2: 假设红色圈后面仍有 layer, 则计算输出层的 C 对 z 的偏导, 然后往前推

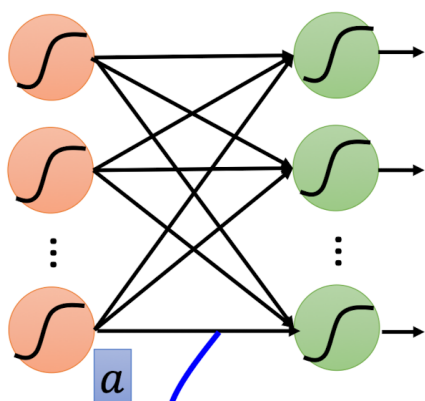


Compute $\partial C / \partial z$ from the output layer

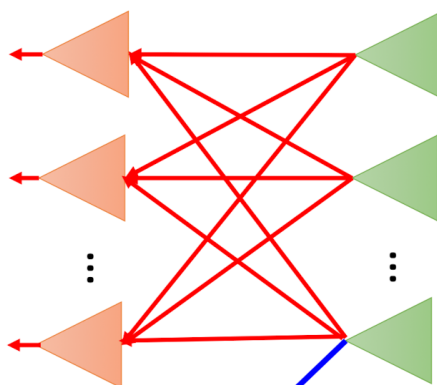


Summary

Forward Pass



Backward Pass



$$\frac{\partial z}{\partial w} = a \quad \times \quad \frac{\partial C}{\partial z} = \frac{\partial C}{\partial w} \quad \Delta$$

forward pass 可以算出每个 activation 的 output (即 $\frac{\partial z}{\partial w}$) for all w

