参数 $\vec{\theta} = [w_1, w_2, w_3, \cdots, b_1, b_2, \cdots]$

$\vec{\theta}^0 = \text{random}()$;

$\vec{\theta}^{t+1} = \vec{\theta}^t - \eta \nabla L(\vec{\theta})$

$\theta^0 \to \theta^1 \to \theta^2 \cdots$

Backpropagation：神经网络通常有成百上千个参数. Back prop 是提供了一个更有效地参数计算技术.

又名 cost function 或 loss function

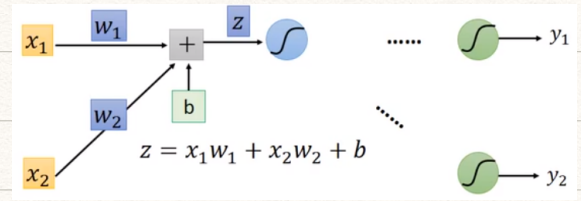
(Cost

定义：$C^n(\vec{\theta})$ 为 $\vec{\theta}$ 为 $\vec{\theta}_j$ 时，input 第 n 维的预测与 label 第 n 维之间的差距.

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

$\Rightarrow \frac{\partial L(\vec{\theta})}{\partial w} = \sum_{i=1}^{n} \frac{\partial l^i(\vec{\theta})}{\partial w}$

取一个 neuron 来看

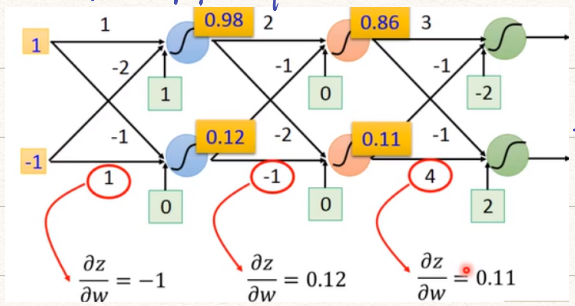


$\frac{\partial l}{\partial w} = \frac{\partial z}{\partial w} \cdot \frac{\partial l}{\partial z}$

Backward Pass：对所有激活函数的输入 z，计算 $\frac{\partial l}{\partial z}$

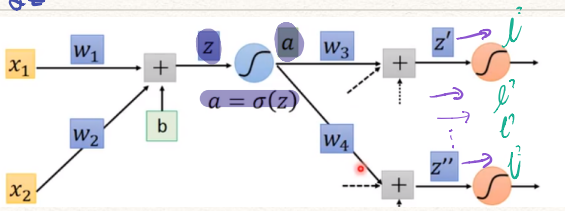
Forward Pass：对所有参数算 $\frac{\partial z}{\partial w}$

对于w而言，计算很简单：$z = w_1 x_1 + w_2 x_2 + b$



$\frac{\partial z}{\partial w_i} = x_i$

$\frac{\partial C}{\partial z}$：



$$\frac{\partial l}{\partial z} = \frac{\partial a}{\partial z} \cdot \frac{\partial l}{\partial a}$$

$$\frac{\partial l}{\partial a} = \frac{\partial l}{\partial z'} \cdot \frac{\partial z'}{\partial a} + \frac{\partial l}{\partial z''} \cdot \frac{\partial z''}{\partial a}$$

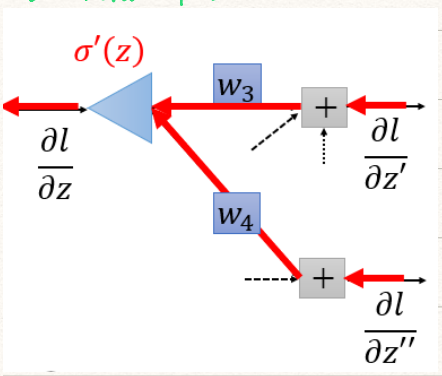
$\frac{\partial z'}{\partial a}$ 为 w_3，$\frac{\partial z''}{\partial a}$ 为 w_4

从后往前连接。

因此有

$$\frac{\partial l}{\partial z} = \sigma'(z)\left(w_3 \cdot \frac{\partial l}{\partial z'} + w_4 \cdot \frac{\partial l}{\partial z''}\right)$$

Backward – Pass：



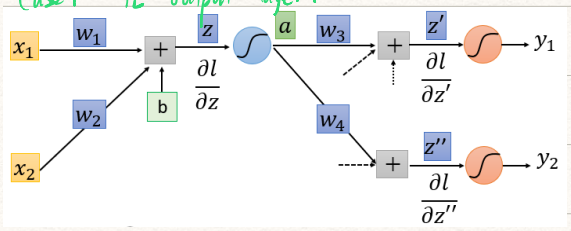
将 $\frac{\partial l}{\partial z'}$ 与 $\frac{\partial l}{\partial z''}$ 看作反向的2个输入

构建与反向的神经网络图符合

在 forward pass 中，$\sigma'(z) = \sigma(z)\frac{1}{1+\sigma(z)}$ 是可算的常数

解决 $\frac{\partial l}{\partial z'}$、$\frac{\partial l}{\partial z''}$

Case 1：在 output layer：

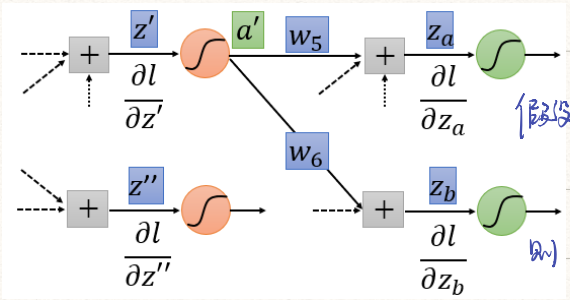


根据你选择的 loss function

$$\frac{\partial l}{\partial z'} = \frac{\partial l}{\partial y_1} \cdot \frac{\partial y_1}{\partial z'} \quad \sigma'(z)$$

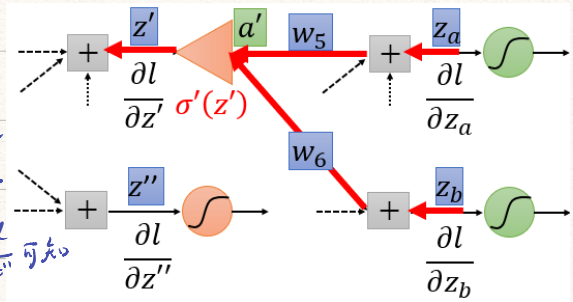
$$\frac{\partial l}{\partial z''} = \frac{\partial l}{\partial y_2} \cdot \frac{\partial y_2}{\partial z''} \quad \sigma'(z'')$$

Case 2：在 Hidden layer：



假设已知 $\frac{\partial l}{\partial z_a}$ 与 $\frac{\partial l}{\partial z_b}$

则 $\frac{\partial l}{\partial z'}$、$\frac{\partial l}{\partial z''}$ 可求



从 output layer 往前算即可