△ 反向传播.

KEYWORDS： Forward Pass . Backward Pass .

train一个神经网络实际是在优化一个loss function.

e.g. 数字识别. input为256-dim vector, output为10-dim vector.

用交叉熵 $C(y,\hat{y}) = -\sum_{i=1}^{10} \hat{y}_i \ln y_i$ 定义loss function：

目标

相当于求 $y, \hat{y}$ 间距离

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

"1"： $y = \begin{bmatrix} y_1 \\ \vdots \\ y_{10} \end{bmatrix}_{10\times 1}$ $\hat{y} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{10\times 1}$. $\theta = \begin{bmatrix} w_1 \\ \vdots \\ b_1 \\ \vdots \end{bmatrix}$ $\nabla L(\theta) = \begin{bmatrix} \frac{\partial L}{\partial w_1} \\ \vdots \\ \frac{\partial L}{\partial b_1} \\ \vdots \end{bmatrix}$

$\theta^* = \arg\min_\theta L(\theta)$.

gradient descent： $\theta^i \leftarrow \theta^{i-1} - \eta \nabla L(\theta^{i-1})$

learning rate.

Backpropagation：算 $\frac{\partial L}{\partial w_1} \cdots \frac{\partial L}{\partial b_1} \cdots$ 的高效方式. （因为参数可能有百万个，直接算太多了）.

$\boxed{\text{Chain Rule}}$：

① $y = g(x)$ $z = h(y)$ $\Delta x \to \Delta y \to \Delta z$

$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx}$ （按路径从后往前算）

② $x = g(s)$ $y = h(s)$ $z = k(x,y)$ $\Delta s \to \Delta x \to \Delta z$, $\Delta s \to \Delta y \to \Delta z$

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

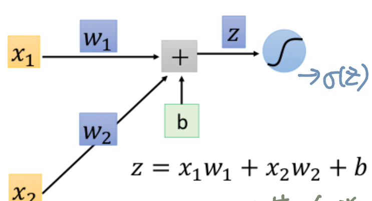
这里是偏导，因为z有2个参数.

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

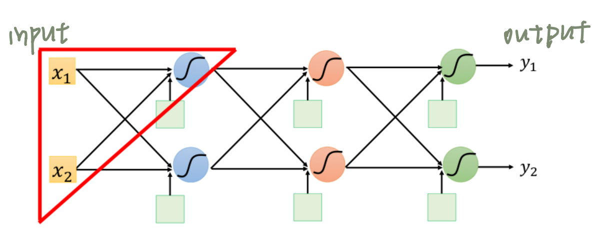
$C(\theta)$ 相当于 $C(y_1, \dots, y_n)$

我们的目标：算C对每个参数的偏微分.

一个neuron：



$z = x_1 w_1 + x_2 w_2 + b$



w为某一参数：

$z = z(w)$ $\Delta w \to \Delta z \to \Delta C$

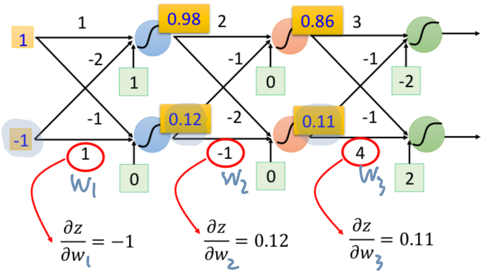
$C = C(z)$ $\therefore \frac{\partial C}{\partial w} = \frac{\partial C}{\partial z}\frac{\partial z}{\partial w}$

难算 好算（有先述过）

- **Forward Pass**：算$\frac{\partial z}{\partial w}$

$\frac{\partial z}{\partial w_1}=x_1$. $\frac{\partial z}{\partial w_2}=x_2$. 即看参数w前连的input是什么，结果就是什么.

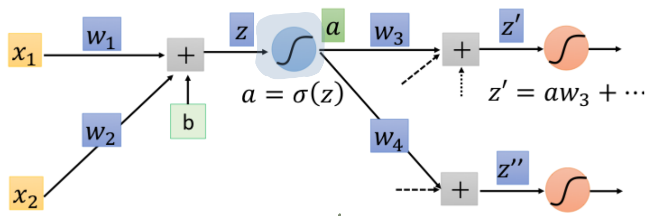
正向计算出所有neuron的output(z)后



∴给定一组input (x)和参数θ，可以直接读出$\frac{\partial z}{\partial w}$.

(z是相对每个neuron来说的).

- **Backward Pass**：算$\frac{\partial C}{\partial z}$



$a=\sigma(z)$
$C=C(a)$
$\Rightarrow \frac{\partial C}{\partial z}=\frac{\partial C}{\partial a}\frac{\partial a}{\partial z}=\sigma'(z)$

$\frac{\partial C}{\partial a}$：难算.

→ 用chain rule表示时，不一定要把中间每个函数都拆出来，因为想拆多少都是能拆出来的. 重要的是能否拆出一部分可求导的.

$\sigma'(z)$：是常数，因为z已被计算出值.

$z'=z'(a)$
$z''=z''(a)$

$C=C(z',z'')$.

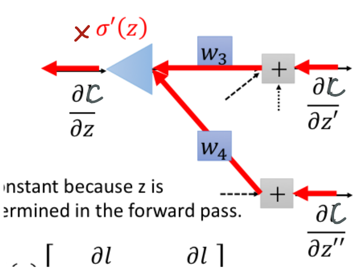
$\Delta a \to \Delta z' \to \Delta C$
$\Delta a \to \Delta z'' \to \Delta C$

$$\frac{\partial C}{\partial a}=\frac{\partial C}{\partial z'}\underbrace{\frac{\partial z'}{\partial a}}_{=w_3}+\frac{\partial C}{\partial z''}\underbrace{\frac{\partial z''}{\partial a}}_{=w_4}$$

$$\therefore \frac{\partial C}{\partial a}=w_3\frac{\partial C}{\partial z'}+w_4\frac{\partial C}{\partial z''}.$$

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

用一个倒过来的neuron表示：



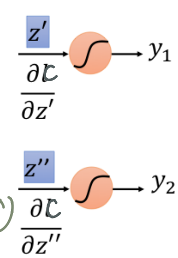
① 当z'、z''不在output layer：

求$\frac{\partial C}{\partial z'}$、$\frac{\partial C}{\partial z''}$和求$\frac{\partial C}{\partial z}$相同.

② 当z'、z''在output layer：

$\left.\begin{array}{l} y_1=\sigma(z') \\ C=C(y_1) \end{array}\right\}$ $\frac{\partial C}{\partial z'}=\frac{\partial C}{\partial y_1}\frac{\partial y_1}{\partial z'}$, $\frac{\partial y_1}{\partial z'}=\sigma'(z')$

同理，$\frac{\partial C}{\partial z''}=\frac{\partial C}{\partial y_2}\frac{\partial y_2}{\partial z''}$, $\frac{\partial y_2}{\partial z''}=\sigma'(z'')$.



$\frac{\partial C}{\partial y}$用C的定义直接求.

▵ Framework

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

$$\frac{\partial C(\theta)}{\partial w}=\boxed{\frac{\partial C}{\partial z}}\frac{\partial z}{\partial w}$$ (z为w所在neuron的线性组合值)

$\frac{\partial z}{\partial w}$ ⟶ Forward Pass，直接读input.

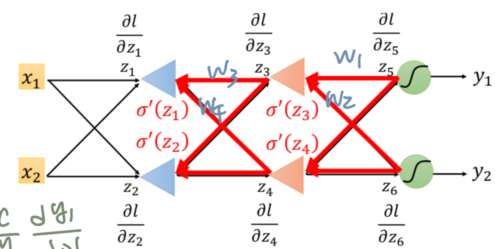
$\frac{\partial C}{\partial z}$ ⟶ Backward Pass

$$\frac{\partial C}{\partial z}=\frac{\partial C}{\partial a}\underbrace{\frac{\partial a}{\partial z}}_{=\sigma'(z)}$$

(z', z''为以σ(z)为input的neuron的线性组合值).

$$\frac{\partial C}{\partial a}=\frac{\partial C}{\partial z'}\underbrace{\frac{\partial z'}{\partial a}}_{=w_3}+\frac{\partial C}{\partial z''}\underbrace{\frac{\partial z''}{\partial a}}_{=w_4}$$

- z', z''在output layer: $\begin{cases}\frac{\partial C}{\partial z'}=\frac{\partial C}{\partial y_1}\frac{\partial y_1}{\partial z'}\\ \frac{\partial C}{\partial z''}=\frac{\partial C}{\partial y_2}\frac{\partial y_2}{\partial z''}\end{cases}$ 均易求.
- z', z''不在output layer: 继续递归



可写出:

$$\frac{\partial l}{\partial z_5}=\frac{\partial l}{\partial y_1}\frac{\partial y_1}{\partial z_5}$$

$$\frac{\partial l}{\partial z_3}=\sigma'(z_3)\left(w_1\frac{\partial l}{\partial z_5}+w_2\frac{\partial l}{\partial z_6}\right)$$

$$\frac{\partial l}{\partial z_1}=\sigma'(z_1)\left(w_3\frac{\partial l}{\partial z_3}+w_4\frac{\partial l}{\partial z_4}\right)$$