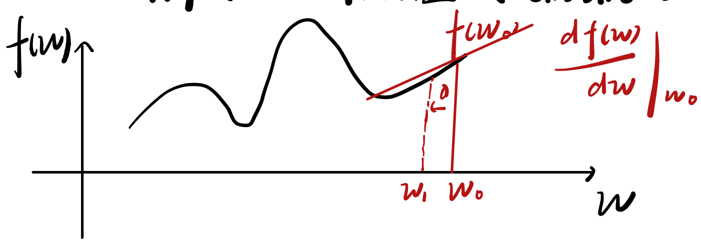
反向传播算法（Back Propogation）

梯度下降法求局部极值（Gradient Descent Method）



① 找一个 w_0

② 设 $k=0$，假设 $\left.\frac{df(w)}{dw}\right|_{w_k}=0$ 退出

否则：$w_{k+1}=w_k-a\left.\frac{df(w)}{dw}\right|_{w_k}$

原因如下：

泰勒展开：

$$f(w+\Delta w)=f(w)+\left.\frac{df(w)}{dw}\right|_{w}\cdot\Delta w+o(\Delta w)$$

$$f(w_{k+1})=f\left(w_k-a\left.\frac{df(w)}{dw}\right|_{w_k}\right)$$

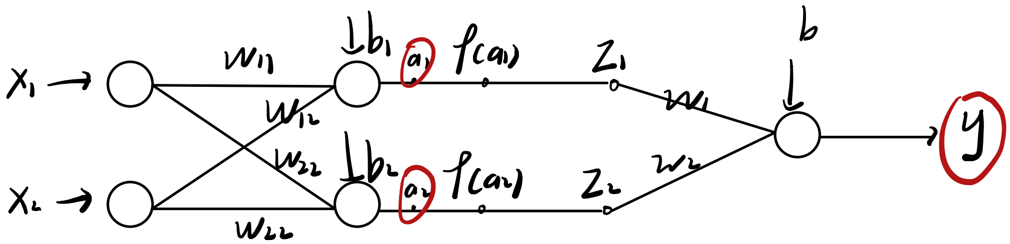
$$=f(w_k)+\left.\frac{df(w)}{dw}\right|_{w_k}\cdot\left(-a\left.\frac{df(w)}{dw}\right|_{w_k}\right)$$

$$+\ o(\Delta w)$$

$$=f(w_k)-\underset{>0}{a}\underset{>0}{\left[\left.\frac{df(w)}{dw}\right|_{w_k}\right]^2}+\underset{\to 0}{o(\Delta w)}$$

$$<f(w_k)$$

BP:



输入: $\{(X_i, Y_i)\}_{i=1\sim N}$

针对输入 (X, Y)

定义 $E = \frac{1}{2}(y-Y)^2$

① 随机取 $(w_{11}, w_{12}, w_{21}, w_{22}, b_1, b_2, w_1, w_2, b)$

② 对所有 w: 求 $\frac{\partial E}{\partial w}$ BP?

b: 求 $\frac{\partial E}{\partial b}$

③ $w^{(新)} = w^{(旧)} - \alpha \frac{\partial E}{\partial b}$

$b^{(新)} = b^{(旧)} - \alpha \frac{\partial E}{\partial b}$

④ 当所有 $\frac{\partial E}{\partial w} / \frac{\partial E}{\partial b}$ 都为0时 退出

$a_1 = w_{11}X_1 + w_{12}X_2 + b_1$

$a_2 = w_{21}X_1 + w_{22}X_2 + b_2$

$Z_1 = f(a_1)$

$Z_2 = f(a_2)$

$y = w_1 Z_1 + w_2 Z_2 + b$

$$\frac{dE}{dy} = (y - Y)$$

$$\frac{\partial E}{\partial a_1} = \frac{dE}{dy} \cdot \frac{\partial y}{\partial Z_1} \cdot \frac{dZ_1}{da_1}$$

$$= (y - Y)\, W_1 \cdot f'(a_1)$$

$$\frac{\partial E}{\partial a_2} = \frac{dE}{dy} \cdot \frac{\partial y}{\partial Z_2} \cdot \frac{dZ_1}{da_2}$$

$$= (y - Y)\, W_2\, f'(a_2)$$

$$\frac{\partial E}{\partial b} = \frac{dE}{dy} \cdot \frac{\partial y}{\partial b} = (y - Y)$$

$$\frac{\partial E}{\partial w_1} = \frac{dE}{dy} \cdot \frac{\partial y}{\partial w_1} = (y - Y)\, Z_1$$

$$\frac{\partial E}{\partial w_2} = \frac{dE}{dy} \cdot \frac{\partial y}{\partial w_2} = y - Y)\, Z_2$$

$$\frac{\partial E}{\partial w_{11}} = \frac{\partial E}{\partial a_1} \cdot \frac{\partial a_1}{\partial w_{11}} = y - Y)\, w_1\, f'(a_1) \cdot X_1$$

$$\frac{\partial E}{\partial W_{12}} = \frac{\partial E}{\partial a_1} \cdot \frac{\partial a_1}{\partial W_{12}} = (y - Y)\, w_1\, f'(a_1)\, X_2$$

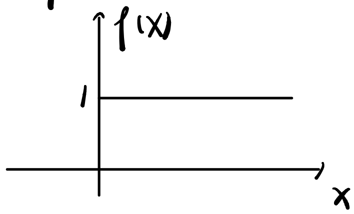
$$\frac{\partial E}{\partial b_1} = (y - Y)\, w_1\, f'(a_1)$$

$$\frac{\partial E}{\partial w_{21}} = (y - Y)\, w_2\, f'(a_2)\, X_1$$

$$\frac{\partial E}{\partial w_{22}} = (y - Y)\, W_2\, f'(a_2)\, X_2$$

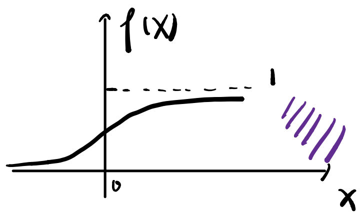
$$\frac{\partial E}{\partial b_2} = (y-\hat{y})\, w_2\, f'(a_2)$$

若 $f(x)$ 取阶跃函数



$$f'(x)=0$$

1.

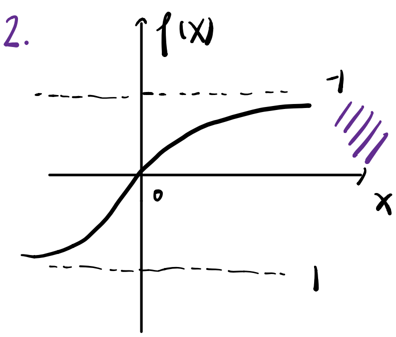


$$f(x)=\frac{1}{1+e^{-x}}$$ (Sigmoid)

信息被压缩

$$f'(x)=f(x)\left[1-f(x)\right]$$

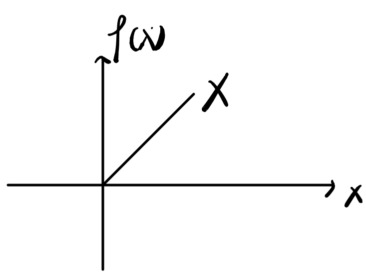
2.



$$f(x)=\tanh(x)=\frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$$

$$f'(x)=1-\left[f(x)\right]^2$$

3.



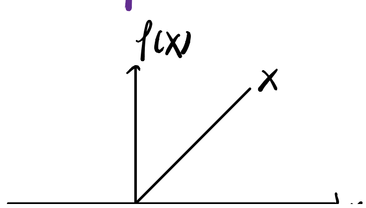
$$f(x)=\begin{cases} x & x>0 \\ 0 & x\le 0 \end{cases} = \max(x,0)$$

$$f'(x)=\begin{cases} 1 & x>0 \\ 0 & x\le 0 \end{cases}$$

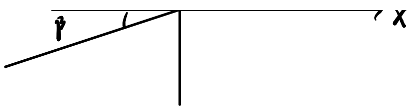
ReLu (x)

Rectified Linear Units

4.



$$f(x)=\begin{cases} x & x>0 \\ \beta x & x\le 0 \end{cases}$$



$$f'(x) = \begin{cases} 1 & x > 0 \\ \beta & x \le 0 \end{cases}$$

Leak Relu(X)