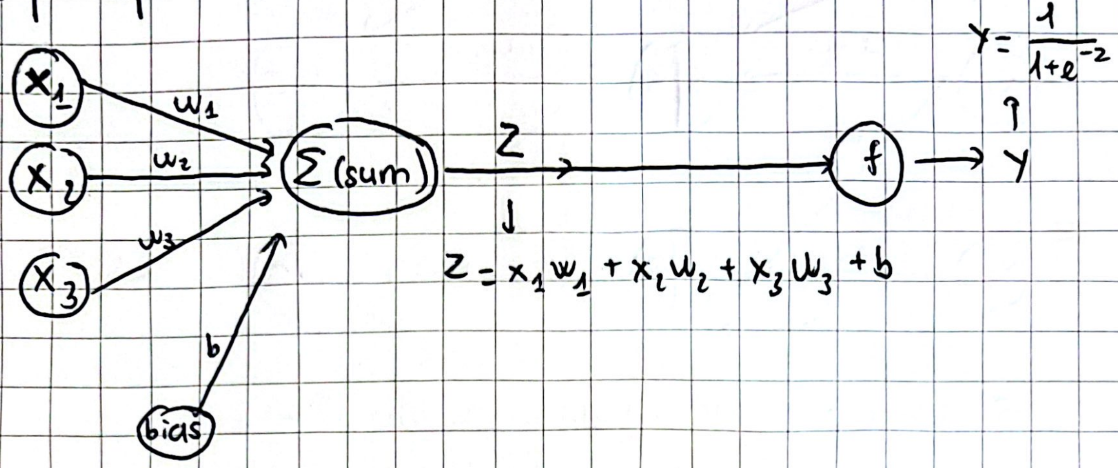
DEEP LEARNING AND NEURAL NETWORK THEORY AND APPLICATIONS - LECTURE NOTE.

Lai Khanh Hoang

- University of Information Technology, Ho Chi Minh City, Viet Nam

Section 1.

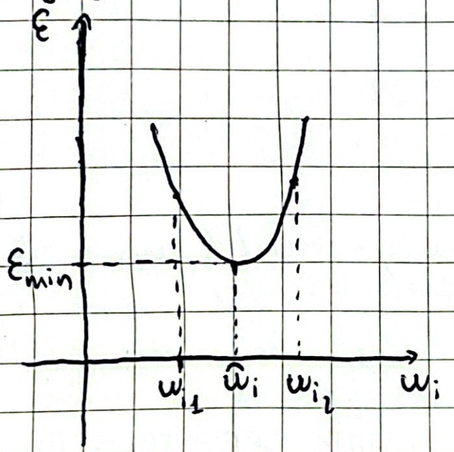
* The perceptron:



* Gradient Descent: an epoch is one complete pass through all the sa

+) Let's consider $-\frac{\partial E}{\partial w_i^k}$, here E is the loss function.

When apply gradient descent, we add this term to the ~~coef~~ weight



Function E decreases in $(; \hat{w}_i)$, so $-\frac{\partial E}{\partial w_{i_1}}$ must be POSITIVE

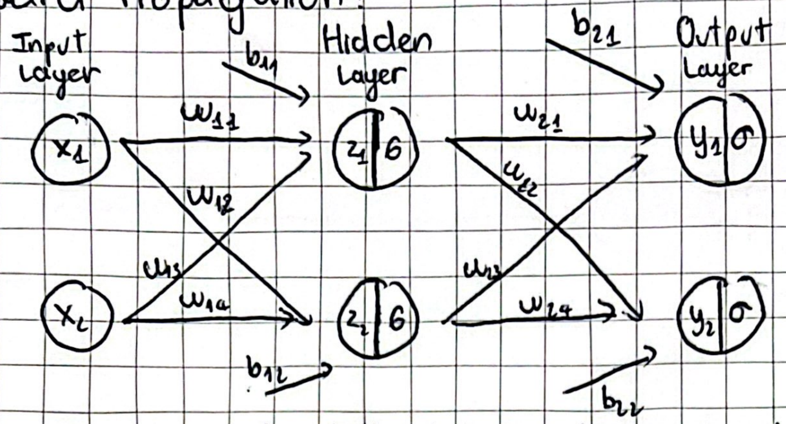
So when we add this term to the weight, it increases

~~Similar to~~ w_{i_2} or $w'_{i_1} = w_{i_1} - \frac{\partial E}{\partial w_{i_1}} > w_{i_1}, \to \hat{w}_i$

Similar to w_{i_2}

+) Weight update Rule: $w \leftarrow w - \eta \frac{dE}{dw}$

* Forward Propagation:



σ is sigmoid-activation Function, $\lambda = 1$.

Input:

Feature 1 x_1	Feature 2 x_2	map to →	Feature 1 d_1	Feature 2 d_2
0.5	-0.5		0.9	0.1
0.3	0.4		0.9	0.9
0.7	0.9		0.9	0.1

Initilization:

$w_{11} = 0.1$, $w_{12} = 0.3$, $w_{13} = -0.2$, $w_{14} = 0.55$, $b_{11} = 0.01$, $b_{12} = -0.02$

$w_{21} = 0.37$, $w_{22} = -0.22$, $w_{23} = 0.9$, $w_{24} = -0.12$, $b_{21} = 0.31$, $b_{22} = 0.27$

Assuming one training sample per iteration (batch size = 1)
First batch,
Forward propagation:

$$z_1 = 0{,}1\,x_1 + x_2\cdot(-0{,}2) + 0.01 = 0{,}5\cdot 0{,}1 + (-0.5)\cdot(-0.2) + 0{,}01$$
$$= 0.16$$

$$z_2 = x_1 \cdot 0.3 + x_2 \cdot 0.55 + -0.02 = 0.5\cdot 0.3 + (-0.5)\cdot 0.55 + (-0.02)$$
$$= -0{,}145$$

$$\sigma(z_1) = \frac{1}{1+e^{-0.16}} = 0.5399, \quad \sigma(z_2) = \frac{1}{1+e^{0.145}} = 0.4638.$$

$$y_1 = 0.5399\cdot 0.37 + 0.4638\cdot 0.9 + 0.31 = 0.9271$$
$$y_2 = 0.5399\cdot(-0.22) + 0.4638\cdot(-0.12) + 0.27 = 0.0955$$
$$\sigma(y_1) = 0.7164, \quad \sigma(y_2) = 0.5238.$$

$$\varepsilon = \frac{1}{2}\left\{\overbrace{[d_1 - \sigma(y_1)]^2}^{\varepsilon_1} + \overbrace{[d_2 - \sigma(y_2)]^2}^{\varepsilon_2}\right\}$$

$$=$$

$$\frac{\partial \varepsilon}{\partial w_{21}} = \frac{d\varepsilon_1}{dw_{21}} = \frac{d\varepsilon_1}{d\sigma(y_1)} \cdot \frac{d(\sigma(y_1))}{dy_1} \cdot \frac{\partial y_1}{\partial w_{21}}$$

$$= -(d_1 - \sigma(y_1)) \cdot \sigma(y_1)[1 - \sigma(y_1)] \cdot \sigma(z_1)$$

$$= -(0{,}9 - 0{,}7164) \cdot 0{,}7164 \cdot [1 - 0{,}7164] \cdot 0{,}5399$$

$$\approx -0{,}0201.$$

$$\Rightarrow w_{21}^{1} = w_{21}^{0} + \eta\left(-\frac{\partial \varepsilon}{\partial w_{21}}\right)$$

$$= 0{,}37 + 1{,}2 \cdot 0{,}0201 = 0{,}3941.$$

$$\frac{\partial \varepsilon}{\partial w_{11}} = \left\{\sum_{J=1}^{2}\left[\frac{\partial \varepsilon}{\partial \sigma(y_J)} \cdot \frac{\partial \sigma(y_J)}{\partial y_J} \cdot \frac{\partial y_J}{\partial(\sigma(z_1))}\right]\right\} \cdot \cancel{\partial z_1}\, \frac{\partial(\sigma(z_1))}{z_1} \cdot \frac{\partial z_1}{\partial w_{11}}$$

$$\cancel{\forall J = 1,\ \frac{\partial \varepsilon}{\partial \sigma(y_1)} \cdot \frac{\partial \sigma(y_1)}{\partial(y_1)} \cdot \frac{\partial(y_1)}{\partial(\sigma(z_1))}} \cdot \frac{\partial(\sigma(z_1))}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_{11}}.$$

$$\frac{\partial(\sigma(z_1))}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_{11}} = \sigma(z_1) \cdot [1 - \sigma(z_1)] \cdot x_1$$

$$= 0.5399 \cdot [1 - 0{,}5399] \cdot 0.5 = 0{,}1242$$

$$\frac{\partial \varepsilon}{\partial(\sigma(y_1))} \cdot \frac{\partial(\sigma(y_1))}{\partial y_1} \cdot \frac{\partial y_1}{\partial(\sigma(z_1))} + \frac{\partial \varepsilon}{\partial(\sigma(y_2))} \cdot \frac{\partial(\sigma(y_2))}{\partial y_2} \cdot \frac{\partial y_2}{\partial(\sigma(z_1))}$$

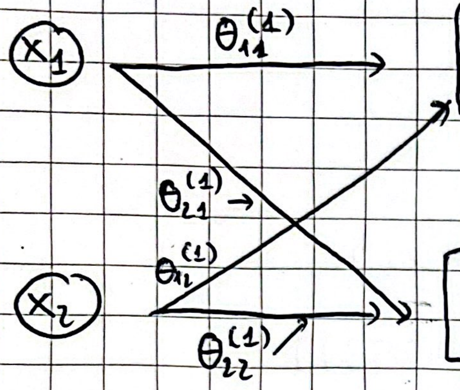
$$= -(0{,}9 - 0{,}7164) \cdot 0{,}7164 \cdot [1 - 0{,}7164) \cdot 0{,}37$$
$$+ -(0{,}1 - 0{,}5238) \cdot \cancel{(1 - 0{,}0955) \cdot 0{,}0955}$$
$$(1 - 0{,}5238) \cdot 0{,}5238) \cdot (-0{,}22)$$

$$\approx -0{,}037$$

So $\dfrac{\partial \varepsilon}{\partial w_{11}} = -0{,}037 \cdot 0{,}2484 \cdot 0{,}5 = -0{,}0045954.$

$$\Rightarrow w_{11} = 0{,}1 + 1{,}2 \cdot 0{,}0045954 = 0{,}1055.$$

Wrap it up:

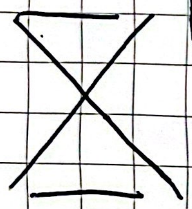


hidden layer

$$\boxed{z_1^{(2)} = \Theta_{11}^{(1)} x_1 + \Theta_{12}^{(1)} x_2 \rightarrow a_1^{(2)} = g(z_1^{(2)})}$$

$$\boxed{z_2^{(2)} = \Theta_{21}^{(1)} x_1 + \Theta_{22}^{(1)} x_2 \rightarrow a_2^{(2)} = g(z_2^{(2)})}$$

Output layer



$$\boxed{z_1^{(3)} = \Theta_{12}^{(2)} a_1^{(2)} + \Theta_{12}^{(2)} a_{2}^{(2)} \rightarrow a_1^{(3)} = g(z_1^{(3)})}$$

output: $a_1^{(3)}$
target: y_1

$$\boxed{z_2^{(3)} = \Theta_{21}^{(2)} a_1^{(2)} + \Theta_{a_{22}}^{(2)} a_2^{(2)} \rightarrow a_2^{(3)} = g(z_2^{(3)})}$$

output: $a_2^{(3)}$
target: y_2.

to

$$X \xrightarrow{\Theta_1} z^{(2)} = \Theta_1 x \rightarrow a^{(2)} = g(z^{(2)}) \xrightarrow{\Theta_2} z^{(3)} = \Theta_2 a^2 \rightarrow a^{(3)} = g(z^{(3)}$$

$$J(\Theta) = \frac{1}{2}(y - a^{(3)})^2.$$