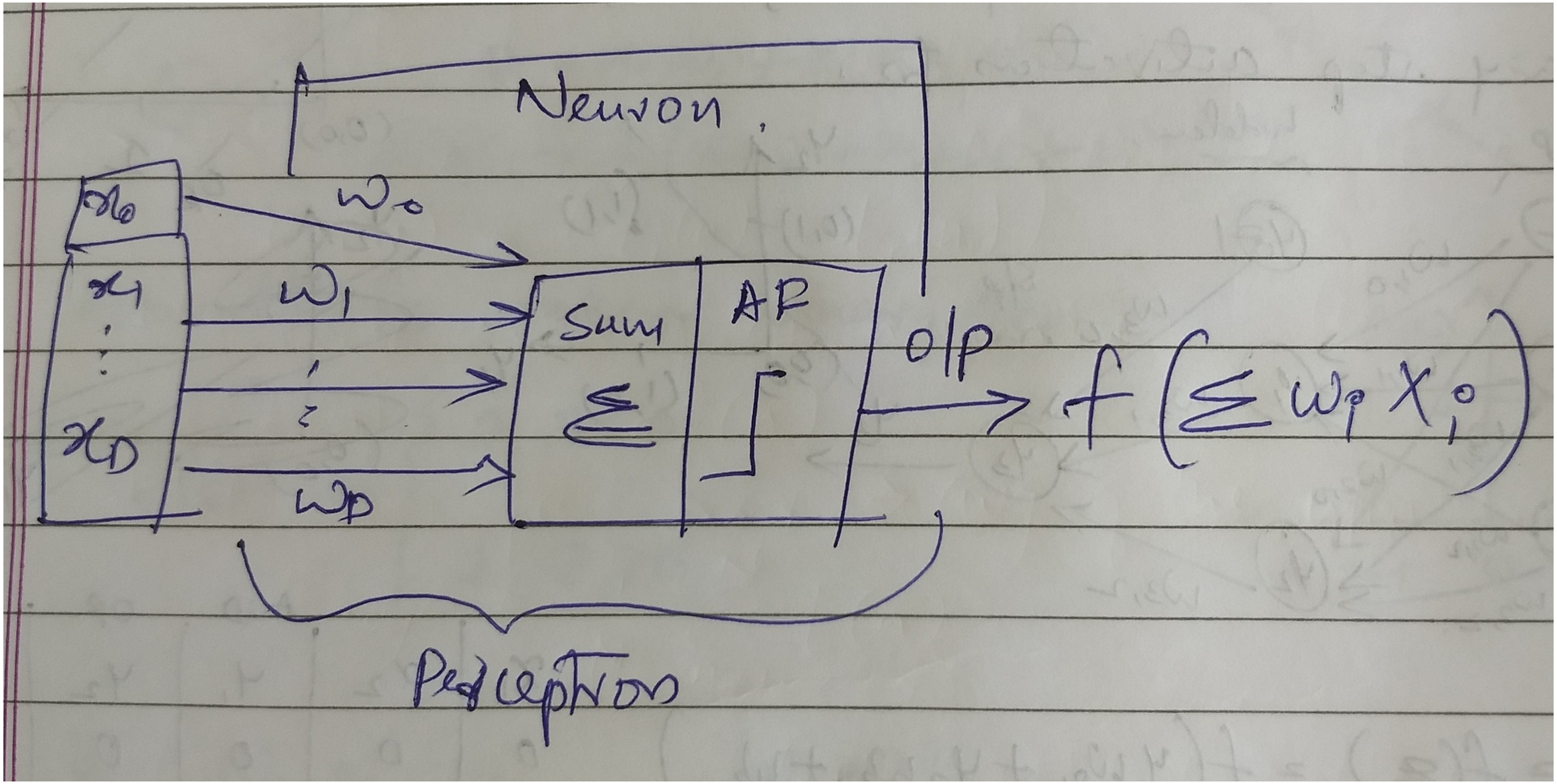
Perceptron, MLP, FP, BP

By, Joe Johnson, D17024

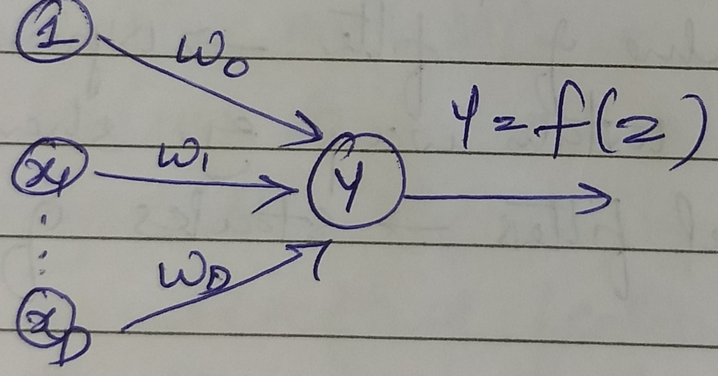
Artificial Neuron/ Perceptron



Propagation rule

Neuron (processing unit).

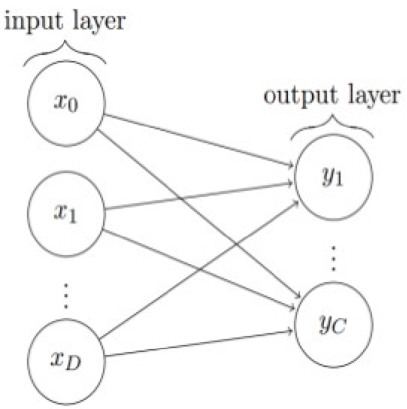
propagation rule → sum of products.



$$z = w_1 x_1 + \dots + w_D x_D + w_0$$

$$f(z) = f\left(\sum_{n=1}^{D} w_n x_n + w_0\right)$$

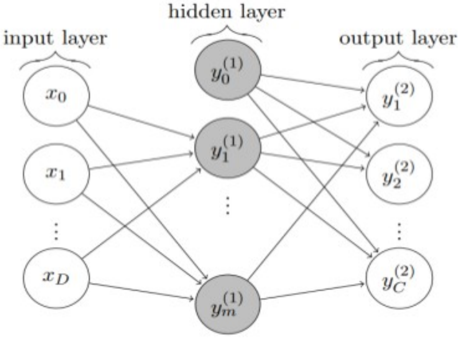
Single layer perceptron (SLP)



- Rosenblatt's perceptron - introduced in 1958
- Perceptron consists of D input units and C output units
  - The additional input value $x_0 := 1$ for biases as weights
  - propagation rule - weighted sum over all inputs with weights and biases
- For $1 \leq i \leq C$ the i^{th} output unit computes the output

$$y_i = f(z_i) \text{ with } z_i = \sum_{k=1}^{D} w_{ik}x_k + w_{i0} \qquad z_i = \sum_{k=0}^{d} w_{ik}x_k$$

Multiple Layer Perceptron(MLP)

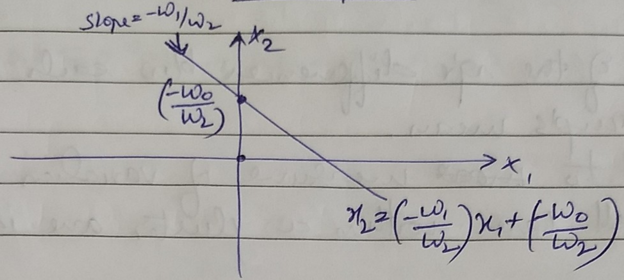


- Additional $L \geq 1$ hidden layers
- l^{th} hidden layer consists of $m^{(l)}$ hidden units

$$y_i^{(l)} = f\left(\sum_{k=1}^{m^{(l-1)}} w_{ik}^{(l)} y_k^{(l-1)} + w_{i0}^{(l)}\right) \overset{y_0^{(l-1)}:=1}{=} f\left(\sum_{k=0}^{m^{(l-1)}} w_{ik}^{(l)} y_k^{(l-1)}\right)$$

$$y(\cdot, w) : \mathbb{R}^D \to \mathbb{R}^C, x \mapsto y(x, w) = \begin{pmatrix} y_1(x, w) \\ \vdots \\ y_C(x, w) \end{pmatrix}$$

Decision Surface / hyperplane.



$X = \{x_1, x_2\} \in \mathbb{R}^2$.

Output of summation block. $z = W^T X + w_0$

$$z = w_1 x_1 + w_2 x_2 + w_0$$

O/p of perceptron: $y = f(z) = f(w_1 x_1 + w_2 x_2 + w_0)$

AF $\rightarrow$ binary ~~step~~ step fn $\Rightarrow$ $f(z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0 \end{cases}$

$\Rightarrow$ $w_1 x_1 + w_2 x_2 + w_0 < 0 \rightarrow f(z) = 0$

$w_1 x_1 + w_2 x_2 + w_0 \geq 0 \rightarrow f(z) = 1$

Condition for decision surface $\Rightarrow$ $w_1 x_1 + w_2 x_2 + w_0 = 0$

$$\Rightarrow x_2 = \underbrace{\left(\frac{-w_1}{w_2}\right)}_{\text{slope of plane}} x_1 + \underbrace{\left(\frac{-w_0}{w_2}\right)}_{\text{intercept on } x_2}$$

$w_n(t+1) = w_n(t) + \eta(y_i - \hat{y}_i)x_i$ until $\rightarrow$ $y_i - \hat{y}_i = 0$

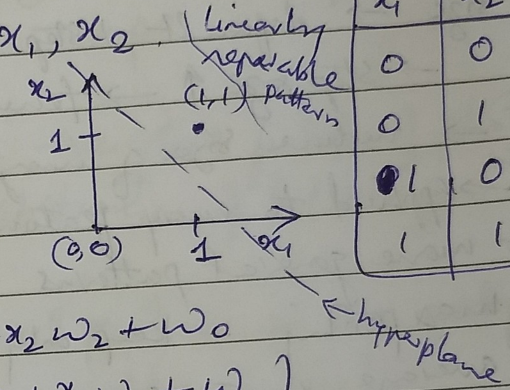
$b(t+1) = b(t) + e$.

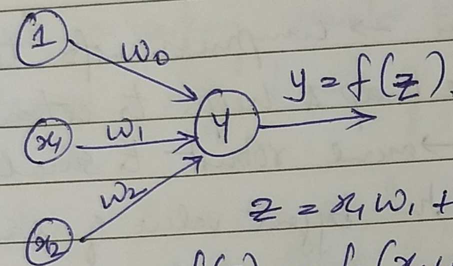
AND Gate

AND

x_1	x_2	y
0	0	0
0	1	0
1	0	0
1	1	1

Consider two input x_1, x_2.

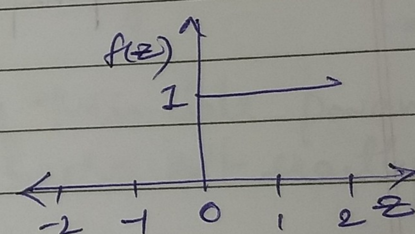




$$z = x_1 w_1 + x_2 w_2 + w_0$$

$$f(z) = f(x_1 w_1 + x_2 w_2 + w_0).$$

Activation fn taken binary step fn



$$f(z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0 \end{cases}$$

let $w_0 = -1.1$, $w_1 = w_2 = 1$.

if, $x_1 = x_2 = 0$

$\Rightarrow z = 0 + 0 - 1.1 = -1.1$

$f(z) = y = 0$

$(w_1 \cdot 0 + w_2 \cdot 0 + w_0 < 0)$
$\Rightarrow w_0 < 0$

if, $x_1 = 1$, $x_2 = 0$

$z = 1 + 0 - 1.1 = -0.1$

$f(z) = 0$ $\quad (w_1 \cdot 1 + w_2 \cdot 0 + w_0 < 0)$
$w_0 < -w_1$

if, $x_1 = 0$, $x_2 = 1 \Rightarrow z = 0 + 1 - 1.1 = -.1$

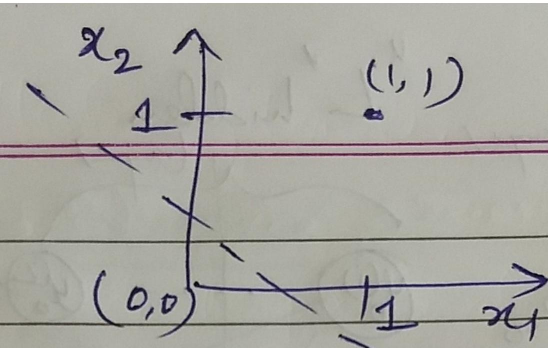
$\Rightarrow y = f(z) = 0$

$(w_1 \cdot 0 + w_2 \cdot 1 + w_0 < 0)$
$\Rightarrow w_0 < -w_2$

if, $x_1 = 1$, $x_2 = 1$

$z = 1 + 1 - 1.1 = 1.9$ [sic: should be 0.9]

$f(z) = 1$ $\quad (w_1 \cdot 1 + w_2 \cdot 1 + w_0 \geq 0)$
$w_1 + w_2 \geq -w_0$



OR Gate

Consider two input, one processing unit, binary step activation fn.

let $w_0 = -0.1$, $w_1 = w_2 = 1$.

x_1	x_2	y	
0	0	0	$w_0 < 0$
0	1	1	$w_2 + w_0 \geq 0$
1	0	1	$w_1 + w_0 \geq 0$
1	1	1	$w_1 + w_2 + w_0 \geq 0$

if $x_1 = x_2 = 0$

$z = 0 + 0 + -0.1 = -0.1 \Rightarrow f(z) = 0 = y$

if $x_1 = 1, x_2 = 0$.

$z = 1 + 0 - 0.1 = 0.9 \Rightarrow f(z) = 1$

if $x_1 = 0, x_2 = 1$

$z = 0 + 1 - 0.1 = 0.9 \Rightarrow y = f(z) = 1$

if $x_1 = x_2 = 1$

$z = 1 + 1 - 0.1 = 1.9 \Rightarrow f(z) = 1$.

Limitation of perceptron :

$w_1.0 + w_2.0 + w_0 < 0 \Rightarrow w_0 < 0$

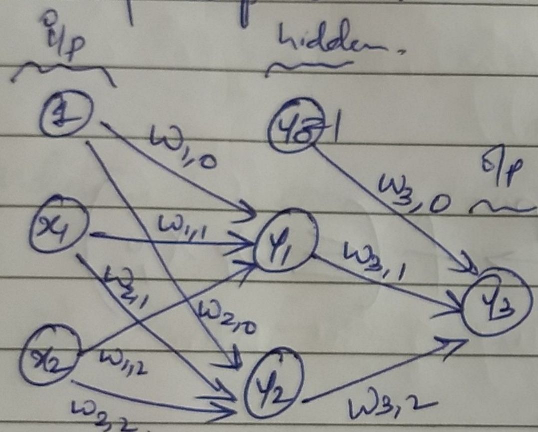
$w_1.0 + w_2.1 + w_0 \geq 0 \Rightarrow w_2 + w_0 \geq 0 \Rightarrow w_2 \geq -w_0$

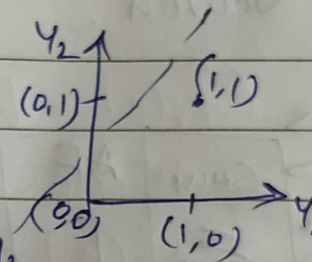
$w_1.1 + w_2.0 + w_0 \geq 0 \Rightarrow w_1 + w_0 \geq 0 \Rightarrow w_1 \geq -w_0$

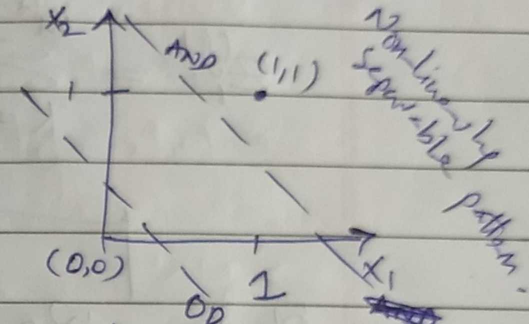
$w_1.1 + w_2.1 + w_0 < 0 \Rightarrow w_1 + w_2 < -w_0$

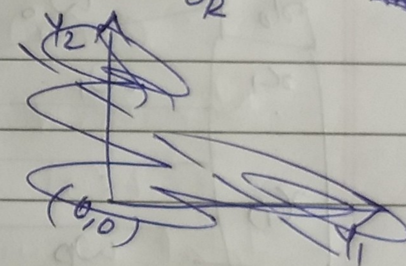
<u>XOR Gate</u>

Consider two input, two processing unit, binary step activation fn.









$$y_3 = f(z) = f(y_1 w_{31} + y_2 w_{32} + w_{30})$$

x_1	x_2	AND y_1	OR y_2	y_3	
0	0	0	0	0	$w_{30} < 0$
0	1	0	1	1	$w_{32} + w_{30} \geq 0$
1	0	0	1	1	$w_{32} + w_{30} \geq 0$
1	1	1	1	0	$w_{31} + w_{32} + w_{30} < 0$

$w_{32} \geq -w_{30}$

$w_{31} < -(w_{30} + w_{32})$

Let $w_{30} = -1.1$, $w_{31} = -1$, $w_{32} = 1.2$

If $y_1 = y_2 = 0$

$\Rightarrow y_3 = f(0 + 0 - 1.1) = 0$

If $y_1 = 0$, $y_2 = 1$

$\Rightarrow y_3 = f(0 + 1.2 - 1.1) = 1$

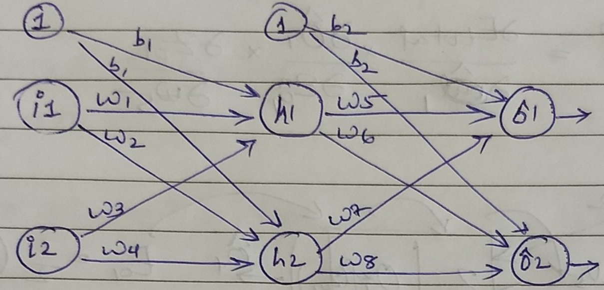
If $y_1 = y_2 = 1 \Rightarrow y_3 = f(-1 + 1.2 - 1.1) = 0$

Training

Rule of thumb: — # of training eg. atleast 5 to 10 times the # of wts.

— $N > \frac{|W|}{(1-a)}$ $\quad |W| \rightarrow$ # of wts

$a \rightarrow$ expected accuracy on test set.



Lets randomly take values for wts & bias b/w 0 & 1.

$w_1 = 0.15, \; w_2 = 0.20, \; w_3 = 0.25, \; w_4 = 0.30, \; b_1 = 0.35$

$w_5 = 0.40, \; w_6 = 0.45, \; w_7 = 0.50, \; w_8 = 0.55, \; b_2 = 0.60$

expected o/p : $o_1 = 0.01, \; o_2 = 0.99$

given i/p : $i_1 = 0.05, \; i_2 = 0.10$

Forward pass: AF $\rightarrow$ (Sigmoid/Logistic)

$$z_{h_1} = w_1 i_1 + w_3 i_2 + b_1 \Rightarrow h_1 = f(z_{h_1})$$

$$h_1 = f(0.15 \times 0.05 + 0.25 \times 0.1 + 0.35) = 1/1+\exp(-(0.0075+0.025+0.35))$$
$$= 1/1+\exp(-0.3825)$$
$$= 1/1+0.6821 = 0.5945$$

$$h_2 = f(w_2 i_1 + w_4 i_2 + b_1) = 1/1+\exp(-0.39) = 1/1+0.6770 = 0.5963$$

$$\hat{o}_1 = f(w_5 h_1 + w_7 h_2 + b_2) = 1/1+\exp(-1.136) = 1/1+0.3211 = 0.7569$$

$$\hat{o}_2 = f(w_6 h_1 + w_8 h_2 + b_2) = 1/1+\exp(-1.195) = 1/1+0.3027 = 0.7676$$

$$E_{total} = \Sigma \frac{1}{2} (\text{target} - \text{output})^2$$

$$= \frac{1}{2}\left[(\overset{o_1}{0.01} - \overset{\hat{o}_1}{0.7569})^2 + (\overset{o_2}{0.99} - \overset{\hat{o}_2}{0.7676})^2\right] = \frac{1}{2}(0.6073) = 0.3037$$

Backwards Pass

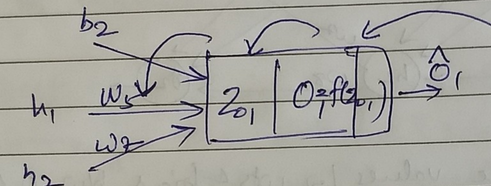
BP → update each of the wts in the N/W → cause the actual o/p to be closest to target o/p → minimizing error.

O/p layer

To check how much change in w_5 affects the total error.

partial derivative or gradient w.r.t w_5 →

$$\frac{\partial E_{total}}{\partial w_5} = \frac{\partial E_{total}}{\partial \hat{O}_1} \times \frac{\partial \hat{O}_1}{\partial z_{o_1}} \times \frac{\partial z_{o_1}}{\partial w_5}$$



$$E_{o_1} = \frac{1}{2}(O_1 - \hat{O}_1)^2$$

$$E_{total} = E_{o_1} + E_{o_2}.$$

Total error change w.r.t to o/p O_1?

$$E_{total} = E_{o_1} + E_{o_2} = \frac{1}{2}(O_1 - \hat{O}_1)^2 + \frac{1}{2}(O_2 - \hat{O}_2)^2$$

$$\frac{\partial E_{total}}{\partial \hat{O}_1} = 2 \cdot \frac{1}{2}(O_1 - \hat{O}_1) \times -1 + 0 = (\hat{O}_1 - O_1)$$
$$= (0.7569 - 0.01)$$
$$= \underline{0.7469}$$

$$\hat{O}_1 = f(z_{o_1}) = \frac{1}{1 + \exp(-z_{o_1})}$$

$$\frac{\partial \hat{O}_1}{\partial z_{o_1}} = \hat{O}_1(1 - \hat{O}_1) = (0.7569)(1 - 0.7569) = \underline{0.1840}$$

$$\frac{\partial z_{o_1}}{\partial w_5} = \frac{\partial (w_5 h_1 + w_7 h_2 + b_2)}{\partial w_5} = h_1 = \underline{0.5945}$$

$$\Rightarrow \frac{\partial E_{total}}{\partial w_5} = 0.7469 \times 0.184 \times 0.5945 = \underline{0.0817}$$

General BP equations

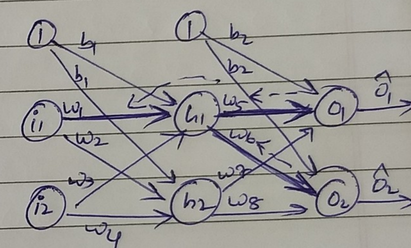
$$\frac{\partial E_{total}}{\partial w_5} = -(O_1-\hat{O}_1) \times \hat{O}_1(1-\hat{O}_1) \times h_1$$
$$= -(O_1-\hat{O}_1) \times f'(z_{O_1}) \times h_1$$
$$= \underline{\underline{\delta \cdot h_1}}$$

To decrease the error,

let $\eta = 0.5$.

updated wt, $w_5^+ = w_5 - \eta * \frac{\partial E_{total}}{\partial w_5} = 0.4 - 0.5 * 0.0817 = \underline{\underline{0.3591}}$

Hidden layer



$$\frac{\partial E_{total}}{\partial w_1} = \frac{\partial E_{total}}{\partial h_1} \times \frac{\partial h_1}{\partial z_{h_1}} \times \frac{\partial z_{h_1}}{\partial w_1}$$

$$\frac{\partial E_{total}}{\partial h_1} = \frac{\partial E_{O1}}{\partial h_1} + \frac{\partial E_{O2}}{\partial h_1}$$

$$\frac{\partial E_{O1}}{\partial h_1} = \frac{\partial E_{O1}}{\partial z_{O1}} \times \frac{\partial z_{O1}}{\partial h_1}, \quad \frac{\partial E_{O1}}{\partial z_{O1}} = \frac{\partial E_{O1}}{\partial \hat{O}_1} \times \frac{\partial \hat{O}_1}{\partial z_{O1}}$$

$$, \frac{\partial z_{O1}}{\partial h_1} = \frac{\partial(w_5 h_1 + w_7 h_2 + b2)}{\partial h_1}$$
$$= w_5$$

Illy,

$$\frac{\partial E_{O2}}{\partial h_2} = \frac{\partial E_{O2}}{\partial z_{O2}} \times \frac{\partial z_{O2}}{\partial h_2}$$

$$\frac{\partial h_1}{\partial z_{h_1}} = \frac{\partial(1/1+\exp(-z_{h_1}))}{\partial z_{h_1}} = h_1(1-h_1)$$

$$\frac{\partial z_{h_1}}{\partial w_1} = \frac{\partial(w_1 i_1 + w_3 i_2 + b_1)}{\partial w_1} = i_1$$

Error Backpropagation

$$\frac{\partial E_n}{\partial w_{ij}^{(L+1)}} \overset{\text{chain-}}{\underset{rule}{=}} \frac{\partial E_n}{\partial z_i^{(L+1)}} \frac{\partial z_i^{(L+1)}}{\partial w_{ij}^{(L+1)}}$$

$$\begin{aligned}\delta_i^{(L+1)} := \frac{\partial E_n}{\partial z_i^{(L+1)}} &\overset{\text{chain-}}{\underset{rule}{=}} \frac{\partial E_n}{\partial y_i^{(L+1)}} \frac{\partial y_i^{(L+1)}}{\partial z_i^{(L+1)}} \\ &= \frac{\partial E_n}{\partial y_i^{(L+1)}} f'\left(z_i^{(L+1)}\right)\end{aligned}$$

$$\begin{aligned}\frac{\partial z_i^{(L+1)}}{\partial w_{ij}^{(L+1)}} &= \frac{\partial}{\partial w_{ij}^{(L+1)}} \left[\sum_{k=0}^{m^{(L)}} w_{ik}^{(L+1)} y_k^{(L)}\right] \\ &= \frac{\partial}{\partial w_{ij}^{(L+1)}} \left[w_{ij}^{(L+1)} y_j^{(L+1)} + \underbrace{\sum_{k=0,k\neq i}^{m^{(L)}} w_{ik}^{(L+1)} y_k^{(L)}}_{const.}\right] \\ &= y_j^{(L)}.\end{aligned}$$

$$\frac{\partial E_n}{\partial w_{ij}^{(L+1)}} = \delta_i^{(L+1)} y_j^{(L)}$$