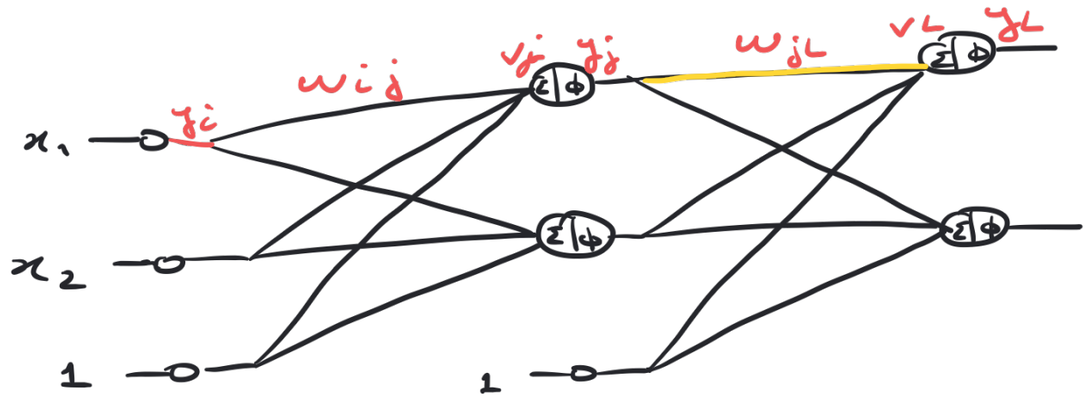


i^{th} layer (input)

j^{th} layer (hidden)

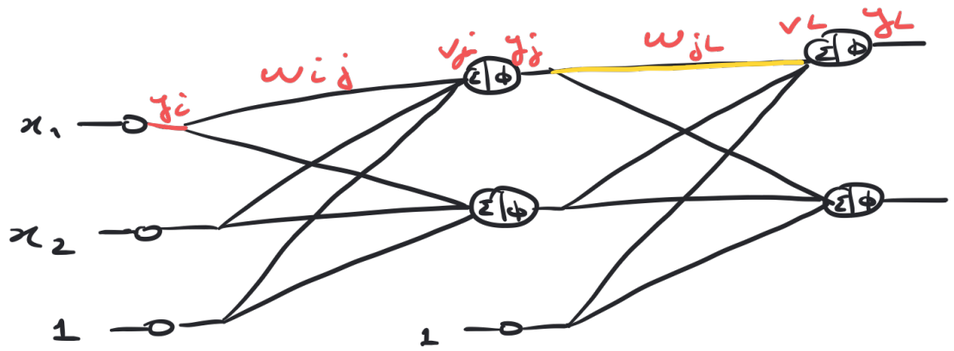
L^{th} layer (output)

OBJECTIVE fct : $J(w) = \frac{1}{2} \sum_{i=1}^{2} \varepsilon_i$

ERROR (at output layer) : $\varepsilon_i = t_i - y_i$

output units : $y_L = \phi\left(\underbrace{\sum_j w_{jL} \cdot y_j}_{v_L} \right) = \phi(v_L)$

weighted sum : $v_L = \sum_j w_{jL} \cdot y_j$



i^{th} layer (input)

j^{th} layer (Hidden)

L^{th} layer (output)

$$\frac{\partial J}{\partial w_{jL}} = \frac{\partial J}{\partial E_L} \cdot \frac{\partial E_L}{\partial y_L} \cdot \frac{\partial y_L}{\partial v_L} \cdot \frac{\partial v_L}{\partial w_{jL}}$$

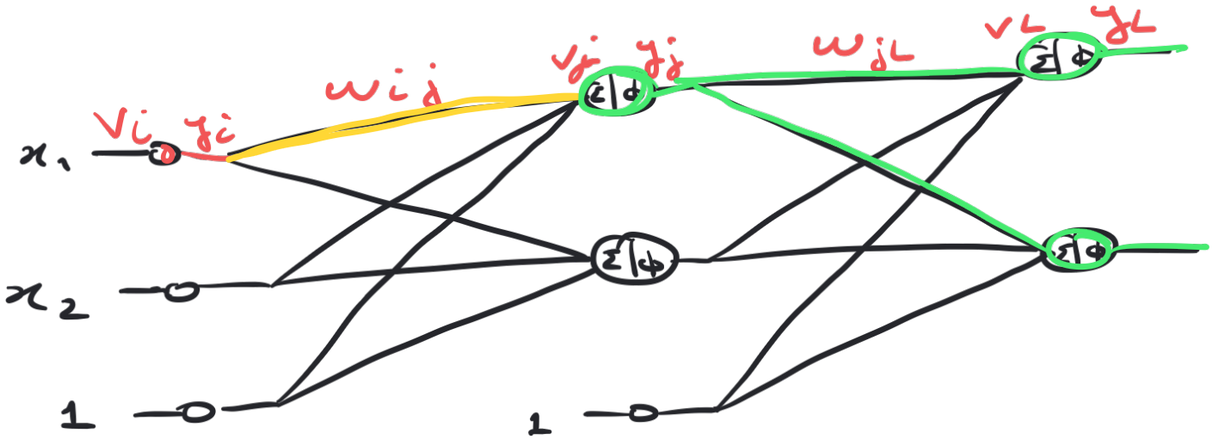
$$= E_L \cdot (-1) \cdot \phi'(v_L) \cdot y_j$$

online learning

it uses a single sample to make corrections

Gradient Descent:

$$w_{jL}^{(t+1)} \longleftarrow w_{jL}^{(t)} + \eta \cdot E_L \cdot \phi'(v_L) \cdot y_j$$



i^{th} layer (input)

j^{th} layer (hidden)

L^{th} layer (output)

δ_i δ_j $\delta_L \leftarrow$ local gradients

they accumulate ERROR seen in the output layer

$$\delta_j = -\frac{\partial J}{\partial v_j}$$

$$\delta_L = -\frac{\partial J}{\partial v_L} \quad \text{and} \quad \delta_i = -\frac{\partial J}{\partial v_i}$$

Looking at the weights at the output layer, w_{jL}:

$$\frac{\partial J}{\partial w_{jL}} = \frac{\partial J}{\partial E_L} \cdot \frac{\partial E_L}{\partial y_L} \cdot \frac{\partial y_L}{\partial v_L} \cdot \frac{\partial v_L}{\partial w_{jL}}$$

$$\underbrace{\qquad\qquad\qquad\qquad}_{-\delta_L}$$

$$\boxed{\frac{\partial J}{\partial w_{jL}} = -\delta_L \cdot \frac{\partial v_L}{\partial w_{jL}} = -\delta_L \cdot y_i}$$

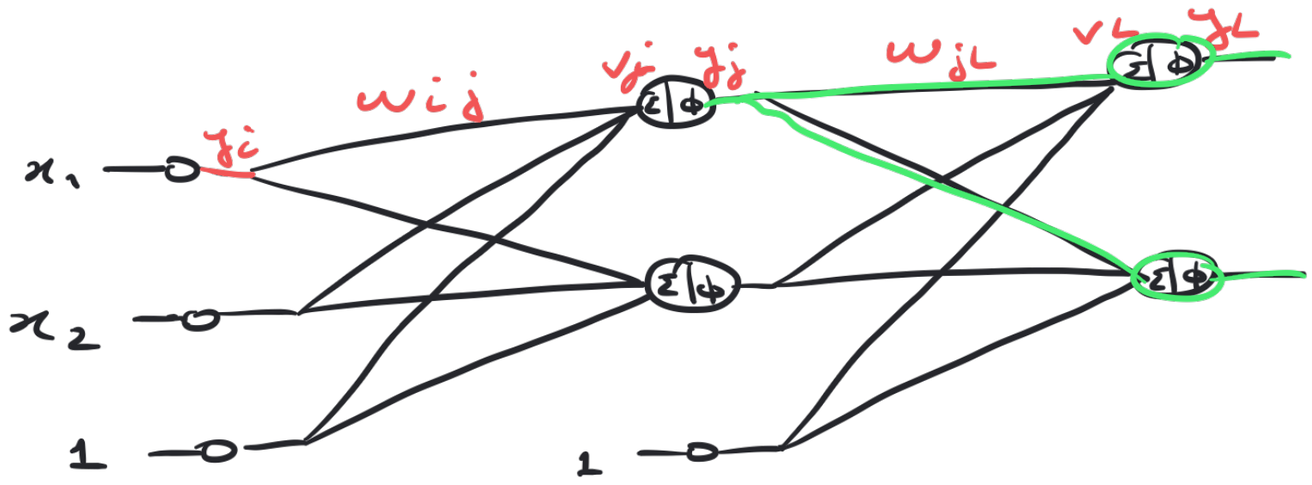
Now, for the weights in the hidden layer, w_{ij}:

$$\frac{\partial J}{\partial w_{ij}} = -\delta_j \cdot y_i$$

$$y_j = \phi(v_j)$$

$$\delta_j = -\frac{\partial J}{\partial v_j} = -\frac{\partial J}{\partial y_j} \cdot \frac{\partial y_j}{\partial v_j}$$

$\phi'(v_j)$



i^{th} layer (input)

j^{th} layer (Hidden)

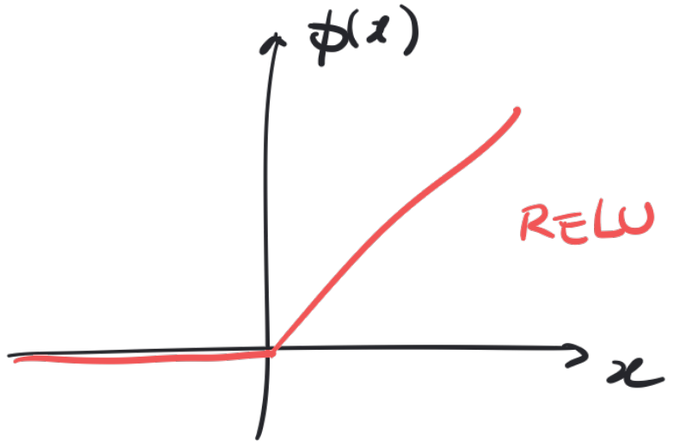
L^{th} layer (output)

$$\frac{\partial J}{\partial y_j} = \sum_L \frac{\partial J}{\partial E_L} \cdot \frac{\partial E_L}{\partial y_L} \cdot \frac{\partial y_L}{\partial v_L} \cdot \frac{\partial v_L}{\partial y_j}$$

$$= \sum_L E_L \cdot (-1) \cdot \phi'(v_L) \cdot w_{jL}$$

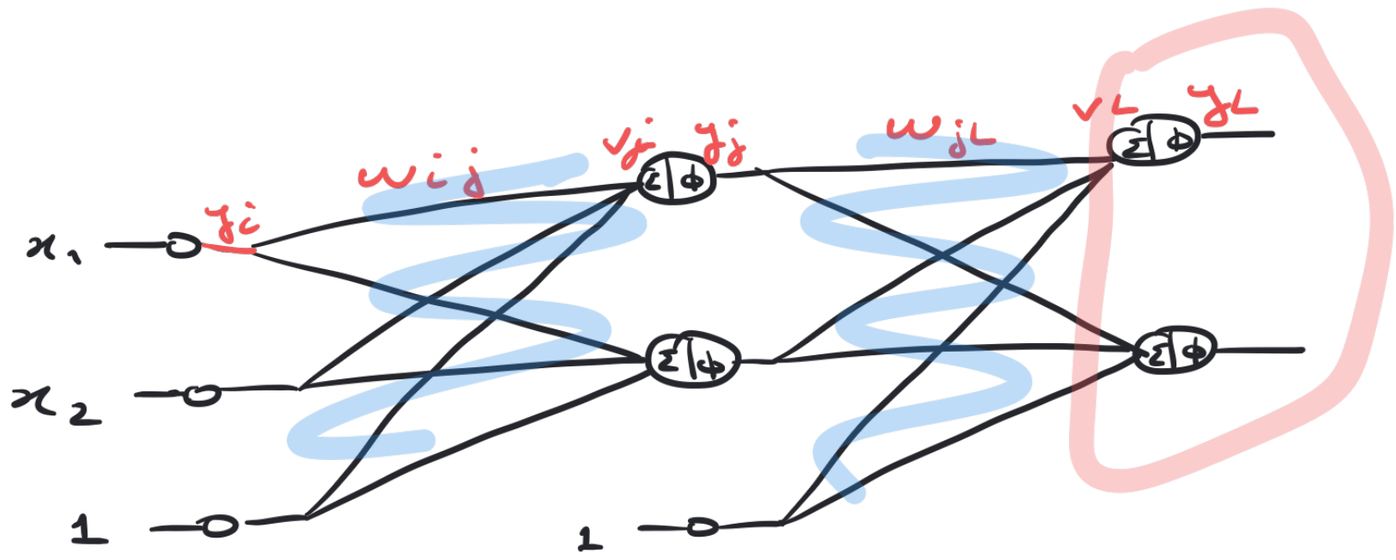
Putting it all together :

$$\delta_j = \left(\sum_L \epsilon_L \cdot \phi'(v_L) \cdot \omega_{jL} \right) \cdot \phi'(v_j)$$



$$\frac{\partial I}{\partial \omega_{ij}} = -\delta_j \cdot y_i$$

$$= -\left(\sum_L \epsilon_L \cdot \phi'(v_L) \cdot \omega_{jL} \right) \cdot \phi'(v_j) \cdot y_i$$



1 0.8
0 0.1
6 0.05
0.05
0

i^{th} layer (input)

j^{th} layer (hidden)

L^{th} layer (output)

class 0

2 − 2 − 2