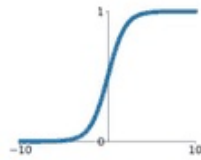


## Activation Functions

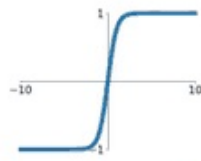
### Sigmoid

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



### tanh

$$\tanh(x)$$



### ReLU

$$\max(0, x)$$

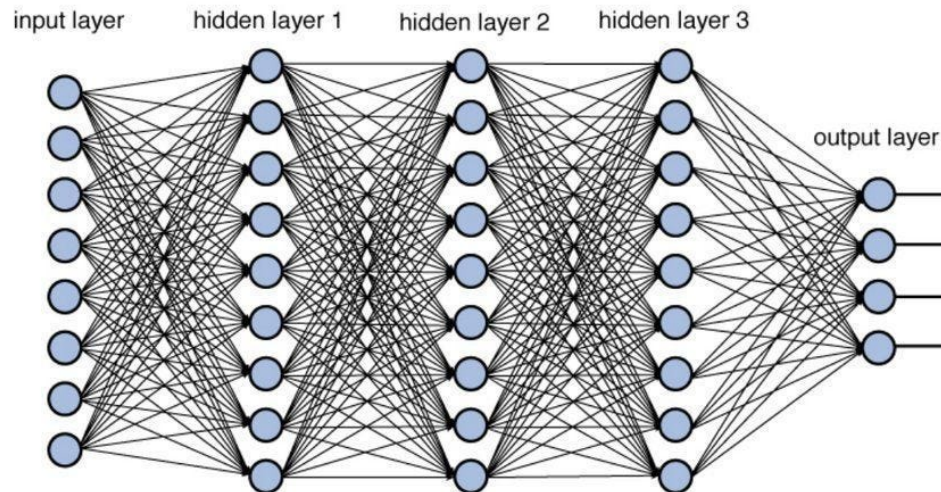
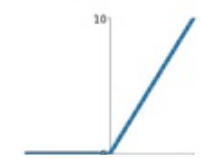


Figure 12.2 Deep network architecture with multiple layers.

## 1 Rétropropagation

$$\delta^1 = (f^1)' \circ (W^2)^T \cdot (f^2)' \circ \dots \circ (W^{L-1})^T \cdot (f^{L-1})' \circ (W^L)^T \cdot (f^L)' \circ$$

$$\delta^2 = (f^2)' \circ \dots \circ (W^{L-1})^T \cdot (f^{L-1})' \circ (W^L)^T \cdot (f^L)' \circ \nabla_{a^L} C$$

$\vdots$

$$\delta^{L-1} = (f^{L-1})' \circ (W^L)^T \cdot (f^L)' \circ \nabla_{a^L} C$$

$$\delta^L = (f^L)' \circ \nabla_{a^L} C,$$

## 2 Algorithme du gradient stochastique

$$W^i = W^i - \lambda \nabla L(W, y, \hat{y})_{W^i} = W^i - \lambda \delta_W^i$$

$$L(W, y, \hat{y}) = \arg(\min_X L(W, y, \hat{y}))$$